Combien de temps durent les situations de monoparentalité ? Une estimation sur données françaises

How long do situations of single parenthood last? An estimation based on French data

Vianney Costemalle *

Résumé – Les familles monoparentales représentent aujourd'hui plus de 20 % des familles avec enfants mineurs en France comme en moyenne en Europe. La monoparentalité est généralement associée à de plus grands risques de précarité et d'exclusion auxquels répondent différentes politiques sociales. Il est ainsi important de savoir combien de temps dure cette situation. On présente dans cet article une méthode originale d'estimation de cette durée à partir d'un échantillon de familles monoparentales pour lesquelles seules les anciennetés dans la situation à la date de l'enquête sont observées (échantillonnage dans le stock). Elle combine un calcul de la fonction de vraisemblance des observations selon la méthodologie proposée par Nickell avec l'introduction de risques instantanés de sortie de la situation proportionnels comme dans le modèle de Cox. Plusieurs simulations reproduisant des cas réels et variés confirment la fiabilité de cette méthode. Appliquée aux données de l'*Enquête famille et logements* de 2011, elle permet d'estimer que la moitié des parents de famille monoparentale sortent de cette situation au bout de trois ans.

***Abstract** – Single-parent families currently account for over 20% of families with minor children in France, in line with the European average. Single parenthood is often associated with greater risks of insecurity and exclusion, to which social policies must respond. It is thus important to know how long such situations last. In this paper, we present an original method for estimating the duration of such periods based on a sample of single parents for whom only the length of time spent in the situation at the time of the survey is observed (stock sampling). It combines a calculation of the likelihood function of the observations using the methodology proposed by Nickell and the introduction of proportional instantaneous probability of exiting the situation based on the Cox model. Several simulations replicating a variety of observed scenarios confirm the reliability of this method. Applying this method to the data from the 2011* Family and Housing Survey *allows us to estimate that single parenthood ends after 3 years for half of the single parents.*

Codes JEL / JEL codes : J12, C14, C15, C41
Mots-clés : monoparentalité, modèle de durée, estimation durées
Keywords : single parenthood, survival models, duration models

** Insee, Division Enquêtes et études démographiques (vianney.costemalle@insee.fr).*

Reçu le 10 octobre 2016, accepté après révisions le 31 mars 2017
DOI: 10.24187/ecostat.2017.493S.1914

La monoparentalité, c'est-à-dire la situation d'un parent – le plus souvent une mère – seul avec un ou des enfants à charge, est une situation familiale dont la fréquence s'est élevée depuis plusieurs décennies[1], en France comme à l'étranger. En France, selon le recensement de la population de l'Insee[2], il y avait 953 000 familles monoparentales en 1990 et 1 687 000 en 2010, ce qui correspond à une progression de 77 % en 20 ans. Parallèlement, le nombre de familles avec enfants mineurs n'a que peu évolué (7 944 000 en 2010, soit 3.8 % de plus qu'en 1990) ; la part des familles monoparentales parmi l'ensemble des familles avec enfants mineurs a ainsi fortement progressé, passant de 12 % en 1990 à 21 % en 2010, dans un mouvement de hausse constante depuis plusieurs dizaines d'années. En Europe, cette tendance a été observée en Grande-Bretagne à partir des années 1970 (David et al., 2004) avant de s'étendre à l'ensemble des pays dans les années 1980. En 2012, la part des familles monoparentales en Europe atteignait 19 %. Les contrastes sont marqués entre pays d'Europe du Nord et pays d'Europe de l'Est et du Sud ; les formes traditionnelles de la famille sont encore bien ancrées dans ces derniers même si la monoparentalité y a fortement progressé (Le Pape et al., 2015).

Plus souvent que d'autres situations familiales, la monoparentalité est associée à des situations difficiles : les familles monoparentales représentent la composition familiale dans laquelle le taux de pauvreté est le plus élevé, atteignant 33 % en France (Boiron et al., 2016), et elles connaissent plus souvent des conditions de logement défavorisées (Chardon et al., 2008) ; les mères de familles monoparentales, bien que leur taux d'activité soit un peu plus élevé que celui des mères en couple, ont un taux de chômage deux fois plus élevé (Rabier, 2014) et font face à des difficultés particulières pour articuler travail et contrainte familiale (Algava et al., 2005). Pour répondre à ces risques d'exclusion professionnelle et sociale, différentes politiques sociales ont été mises en œuvre en Europe dès les années 1950 (Eydoux & Letablier, 2009). En France, l'allocation de soutien familial (ASF) et le complément familial (CF) s'ajoutent ainsi aux divers dispositifs dont bénéficient les familles. Il est donc important de savoir combien de temps durent les épisodes de monoparentalité.

La monoparentalité est une situation forcément transitoire, car elle prend fin soit avec la (re)mise en couple du parent seul, soit lorsque les enfants atteignent un âge auquel ils ne sont plus considérés comme dépendants ou quittent le foyer. Mais la durée des situations de monoparentalité reste à ce jour assez mal connue. L'objectif de cet article est de fournir une estimation de la durée de la monoparentalité, ce qui, à notre connaissance, n'a jamais été fait auparavant pour la France. Pour cela on propose une méthode originale d'estimation de durées à partir des anciennetés, basée sur des travaux anciens de modélisation de durées, notamment les travaux de Nickell (1979) et de Cox (1972).

Avant d'aller plus loin, il faut noter qu'il n'existe pas de définition unique de la monoparentalité. Toutes les approches considèrent néanmoins qu'il s'agit d'un parent isolé élevant des enfants à charge, mais plusieurs critères peuvent être retenus pour chacune de ces deux dimensions. Longtemps, le critère d'isolement fut basé sur le statut matrimonial légal du parent : étaient considérés comme parents isolés les femmes (ou les hommes) non mariés avec des enfants. Cette approche pose maintenant problème puisque de plus en plus de couples ne sont pas mariés. En outre, dans certains couples, les partenaires ne vivent pas dans un même logement, faisant alors partie de ce qu'on dénomme les LAT (*living apart together*). Par exemple, 10 % des parents de familles monoparentales en France déclaraient, en 1999, vivre en couple avec quelqu'un qui n'habite pas dans le même logement (Algava, 2002). Le statut marital n'est donc plus un bon indicateur d'isolement et la notion de « vie en couple » est devenue plus difficile à saisir (Toulemon, 2011). Par ailleurs, il y a aussi plusieurs possibilités pour définir un enfant à charge : le plus souvent, on retient les enfants de moins de 25 ans (par exemple Algava, 2002) ou de moins de 18 ans (Buisson et al., 2005) mais certains travaux ne retiennent pas de limite d'âge (David et al., 2004). L'Insee analyse en général les types de familles en considérant les enfants mineurs ; le Haut Conseil de la famille s'intéresse aussi aux familles avec enfants mineurs (Haut Conseil de la famille, 2014). Nous retiendrons également cette limite d'âge de 18 ans, et nous définissons pour cet article la monoparentalité comme la situation d'un parent ne vivant pas en couple ayant au moins un enfant célibataire âgé de moins de 18 ans vivant dans le même logement, qui n'a pas lui-même d'enfants vivant dans le logement.

La première partie de l'article présente la méthode d'estimation. Dans un premier temps on explique en quoi les anciennetés observées sur

1. Les familles monoparentales apparaissent comme une catégorie statistique en France avec le recensement de 1982.
2. Il s'agit des familles monoparentales avec enfant(s) de moins de 18 ans.

les personnes présentes dans le stock de familles monoparentales sont différentes des durées : les anciennetés peuvent être vues comme des durées doublement biaisées. On détaille ensuite comment il est possible d'inférer la distribution des durées à partir de celle des anciennetés à partir de la connaissance des flux d'entrée. On s'appuie ici sur des simulations numériques, générant des variables aléatoires selon le modèle présenté, pour illustrer différents cas particuliers et pour tester la robustesse de la méthode proposée. Dans la seconde partie, cette méthode est appliquée aux données de l'*Enquête famille et logements* (*EFL*) en utilisant de plus les données de l'enquête *Étude des relations familiales et intergénérationnelles* (*Erfi*) (voir l'encadré 1 pour une présentation de ces sources) ; on expose alors l'estimation de la distribution des durées passées en situation de monoparentalité.

Comment déterminer les durées à partir des anciennetés ?

La durée des situations de monoparentalité est mal connue car difficile à mesurer. Les périodes de monoparentalité se déduisent en effet de la confrontation entre les périodes où les personnes vivent sans conjoint et les périodes où elles ont des enfants à charge. Connaître les dates de début et de fin de ces périodes requiert de disposer d'informations longitudinales (ou rétrospectives) et de telles données sont assez rarement disponibles. Des durées moyennes ont pu être ainsi mesurées aux États-Unis sur la base de données rétrospectives à partir de l'enquête *National Survey of Families and Households* de 1987 : en moyenne, les mères entrées en famille monoparentale entre 1970 et 1974 sont restées 4.5 ans dans cette situation, contre 3.4 ans pour celles entrées 10 ans plus tard (Bumpass & Raley, 1995). Au Royaume-Uni, l'enquête rétrospective *Survey of Family and Working Lives* de 1994 a permis d'estimer des durées médianes de monoparentalité : 5.8 ans pour l'ensemble des mères, 4.6 ans pour les mères célibataires, 4.7 ans pour les mères divorcées, 6.8 ans pour les mères séparées et 10.5 ans pour les mères veuves (McKay, 2002).

Les autres enquêtes plus classiques qui apportent des informations sur les familles monoparentales sont les enquêtes transversales,

Encadré 1

L'*ENQUÊTE FAMILLE ET LOGEMENTS* ET L'ENQUÊTE *ÉTUDE DES RELATIONS FAMILIALES ET INTERGÉNÉRATIONNELLES*

L'*Enquête famille et logements* (*EFL*) a été réalisée par l'Insee au début de l'année 2011 conjointement avec le recensement de la population. Dans chaque ménage sélectionné pour l'enquête, tous les hommes ou toutes les femmes de 18 ans ou plus ont été interrogés. Au total, 359 770 personnes de 18 ans ou plus, vivant en ménage ordinaire en France métropolitaine ont été interrogées sur leur vie familiale et résidentielle (Breuil et al., 2016). Cette enquête permet d'identifier les adultes à la tête d'une famille monoparentale au 1er janvier 2011. Il s'agit donc d'un échantillonnage dans le stock. Les personnes qui déclaraient ne pas vivre en couple au moment de l'enquête, et qui avaient déjà vécu en couple auparavant, devaient indiquer l'année de la fin de l'union ainsi que la cause (séparation ou décès du conjoint). Cela permet notamment de déterminer l'ancienneté de la monoparentalité au moment de l'enquête ainsi que la cause (rupture d'union, enfant hors vie en couple, décès du conjoint) de la monoparentalité.

L'enquête *Étude des relations familiales et intergénérationnelles* (*Erfi*) est le volet français du programme international *Generations and Gender Survey* (*GGS*) visant à établir des statistiques comparatives par pays au niveau mondial, et principalement au niveau européen (Régnier-Loilier, 2012). Elle a été menée conjointement par l'Ined et l'Insee en trois vagues successives, en 2005, 2008 et 2011. Lors de la première vague il y avait 10 079 répondants, puis 6 534 pour la deuxième vague et enfin 5 781 pour la dernière vague (certains ont répondu à la première et troisième vague sans répondre à la deuxième). Les répondants sont âgés de 18 à 79 ans au 31 décembre 2005, une seule personne par ménage étant tirée au sort. L'intérêt de cette enquête est qu'elle pose des questions à caractère rétrospectif, ce qui permet d'avoir des informations longitudinales sur les parcours familiaux et conjugaux. Les personnes interrogées décrivent l'ensemble des relations cohabitantes (dont la durée a été d'au moins trois mois) qu'elles ont vécues au cours de leur vie. Elles donnent aussi des informations sur tous les enfants qu'elles ont eus, avec notamment une information sur la date de départ du foyer de chaque enfant. On a pu, à partir de ces éléments, déterminer les périodes pendant lesquelles les personnes interrogées étaient à la tête d'une famille monoparentale avec enfants mineurs. Les inconvénients de cette enquête sont : une taille de l'échantillon des personnes ayant vécu au moins une fois en famille monoparentale assez faible et de possibles imprécisions sur les dates données, – les questions portant sur un passé parfois lointain – ce qui conduirait à une imprécision sur les épisodes de monoparentalité.

qui permettent de connaître le stock de ces familles à un instant donné. On peut sur cette base mesurer l'ancienneté – mais pas la durée – des situations de monoparentalité. En France, l'enquête *Étude de l'histoire familiale* avait ainsi permis de mesurer pour 1999 une ancienneté moyenne de la situation de 6 ans et 3 mois pour les femmes et 5 ans et 9 mois pour les hommes (Algava, 2002). L'*Enquête famille et logements* de 2011 a permis de mesurer que les 1.5 million de familles monoparentales de France métropolitaine s'étaient formées 5.5 années plus tôt en moyenne : 4.5 ans pour un parents séparé, 5.5 ans pour un parent dont le conjoint est décédé et 10 ans pour un parent n'ayant jamais vécu en couple (Buisson et al., 2015).

Les deux types d'approche, longitudinale et transversale, renvoient respectivement à ce qu'on appelle dans les modèles de durée l'échantillonnage dans le flux (*flow sampling*) et l'échantillonnage dans le stock (*stock sampling*).

Des anciennetés aux durées : le biais de censure et le biais de sélection

L'ancienneté d'une situation est le temps qui s'est écoulé entre le début de cette situation et le moment où la situation est observée[3], tandis que la durée est le temps total écoulé entre le début et la fin de la situation (voir l'encadré 2 pour la modélisation des durées). Les anciennetés et les durées sont donc *a priori* deux concepts différents qui répondent respectivement aux questions « Depuis combien de temps est-on dans cette situation, à la date *t* ? » et « Combien de temps cette situation dure-t-elle ? ». L'ancienneté peut être vue comme une durée qui est censurée sur la droite, puisqu'on ne connaît pas la date de fin de la situation, mais seulement son commencement. Les anciennetés sont par conséquent plus courtes que les durées, ce

3. Le moment de l'observation sera, dans notre cas d'application, identique pour tous les individus : il s'agit de la date de l'enquête.

Encadré 2

MODÉLISATION DES DURÉES

Dans le cas où la variable aléatoire de durée T est discrète et prend des valeurs entières, la loi de T est donné par sa densité $f(t) = P(T = t)$.

La survie au temps $t \in \mathbb{N}$, notée $S(t)$, correspond à la proportion de la population dont la situation d'intérêt a duré t unité de temps ou plus : $S(t) = P(T \geq t)$. Il en résulte que S est une fonction décroissante et que $S(0) = 1$.

Le risque instantané au temps $t \in \mathbb{N}$, noté $h(t)$, correspond à la part de personnes qui quittent la situation au temps t parmi les personnes qui étaient toujours dans la situation au temps t : $h(t) = P(T = t \mid T \geq t)$. Soit $f(t) = h(t)S(t)$.

On a alors une relation entre la survie et le risque instantané : pour tout $t > 0$, $S(t) = \prod_{u=0}^{t-1}(1 - h(u))$.

En remarquant que $f(t) = S(t) - S(t+1)$, on a déduit que l'espérance de la variable de durée est donnée par $E[T] = \sum_{t \geq 0} t f(t) = \sum_{t \geq 1} S(t)$.

La médiane est ici définie comme $Med(T) = \frac{u(1 - S(u) - S(u+1)) + 0,5 - S(u)}{S(u) - S(u+1)}$, où u est le plus grand entier tel que $S(u) \geq 0,5$. Ce qui donne bien $Med(T) = u$ lorsque $S(u) = 0,5$.

Connaître la loi de probabilité, la fonction de survie ou la fonction de risque instantané revient au même et donne toute l'information sur la distribution des durées. Il suffit donc de connaître une de ces trois fonctions pour calculer n'importe quel indicateur, comme la moyenne ou la médiane.

Dans le cas continu, on utilise la loi de Weibull pour simuler des variables de durées. Dans ce cas, on peut définir de la même façon que dans le cas discret la fonction de survie et la fonction de risque instantané. On a alors pour tout $t \in \mathbb{R}$, $S(t) = exp\left(-\int_0^t h(u)du\right)$.

La loi de Weibull est paramétrée par deux réels positifs : un paramètre d'échelle λ et un paramètre de forme k.

La survie s'écrit $S(t) = \exp\left(-\left(\frac{t}{\lambda}\right)^k\right)$ et le risque instantané $h(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1}$.

Lorsque le paramètre de forme k vaut 1, alors on obtient une loi exponentielle : la survie est exponentiellement décroissante et le risque instantané est constant et vaut $1/\lambda$. Si k est plus petit que 1, le risque instantané est décroissant tandis que s'il est plus grand que 1 le risque instantané augmente avec le temps t. De plus, la variance augmente lorsque k diminue.

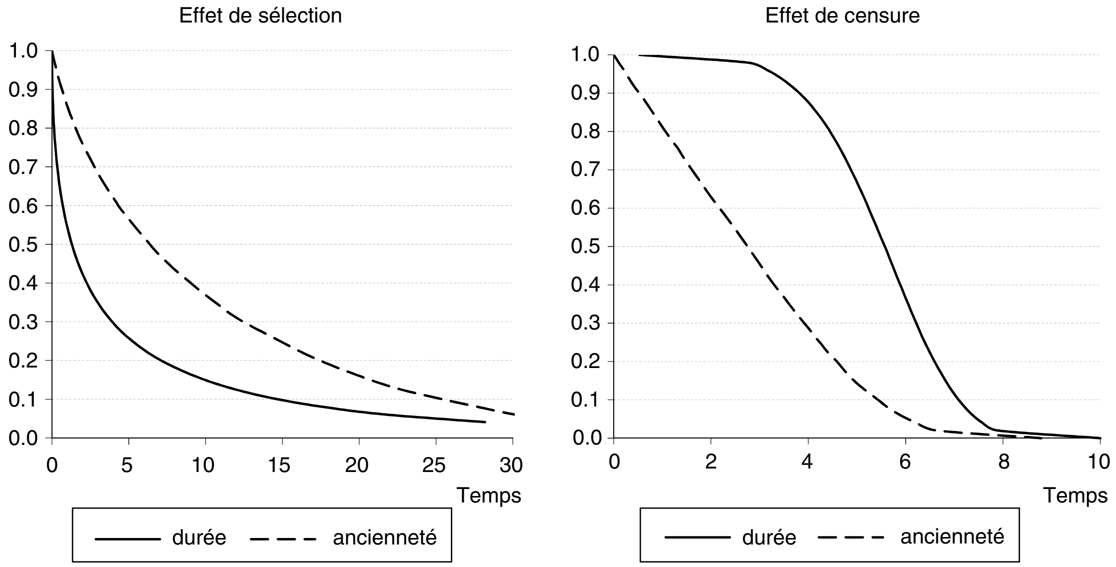
qu'on nomme ici le biais de censure. Un autre problème apparaît dans le cas d'un échantillonnage dans le stock : la probabilité d'être enquêté alors qu'on est en situation monoparentale augmente avec la durée de la monoparentalité. Ceci implique que les personnes en situation monoparentale au moment de l'enquête ont des durées en moyenne plus grandes que l'ensemble des personnes ayant connu un épisode de monoparentalité au cours de leur vie[4]. On nomme cela le biais de sélection. Ce biais joue en sens contraire du biais de censure, ce qui fait qu'il est *a priori* impossible de dire si les anciennetés sont en moyenne plus longues ou plus courtes que les durées. Quel est le lien entre la distribution des durées et celle des anciennetés ? Afin de bien comprendre comment peuvent jouer ces effets de censure et de sélection, on va par la suite générer de façon aléatoire plusieurs jeux de données mimant différentes situations réelles. On pourra alors ajuster les paramètres de ces simulations à volonté pour mettre en évidence les deux biais dans des cas choisis.

On illustre ces phénomènes de censure et de sélection à l'aide de simulations qui génèrent de manière aléatoire des observations de durées et d'anciennetés. On génère aléatoirement, selon une loi uniforme, une date d'entrée en monoparentalité, entre 1950 et 2010, pour 100 000 personnes. On génère de plus pour chacune de ces personnes la variable aléatoire de durée selon une loi de Weibull (cf. encadré 2). On en déduit alors un stock de personnes qui sont toujours en situation de monoparentalité en 2011. Parmi les personnes de ce stock, on peut calculer l'ancienneté de la monoparentalité en 2011. Le paramètre de forme de la loi de Weibull est le paramètre clé : s'il est plus petit que 1, alors c'est l'effet de sélection qui l'emporte tandis que s'il est plus grand que 1 c'est l'effet de censure qui est le plus fort (figure I).

Ce résultat peut se comprendre intuitivement. Dans le cas où le paramètre de forme est inférieur à 1, les durées de la plupart des personnes sont très courtes et seulement quelques-unes d'entre elles ont des durées très longues (il y a une très grande hétérogénéité des durées). Il en résulte qu'au moment de l'enquête, toutes les personnes avec des durées très longues sont en famille monoparentale, tandis que seulement une fraction de celles qui ont des durées très courtes le sont (les personnes qui sont entrées en monoparentalité juste avant l'enquête). Les personnes avec des durées longues sont donc surreprésentées parmi les familles monoparentales au moment de l'enquête. En conséquence, les anciennetés observées sont globalement plus longues que les durées réelles.

4. Pour bien comprendre cela on peut aussi penser au jeu de la bataille navale : en choisissant une case au hasard, les gros bateaux ont plus de chance d'être touchés que les petits.

Figure I
Courbes de survie associées à la distribution des durées et à la distribution des anciennetés illustrant l'effet de sélection et l'effet de censure à partir de simulations



Note : données simulées. Pour la figure de gauche, la durée est générée selon une loi de Weibull(0.5 , 2.75) et pour la figure de droite selon une loi de Weibull(5 , 6). La date d'entrée en monoparentalité est générée selon une loi uniforme.

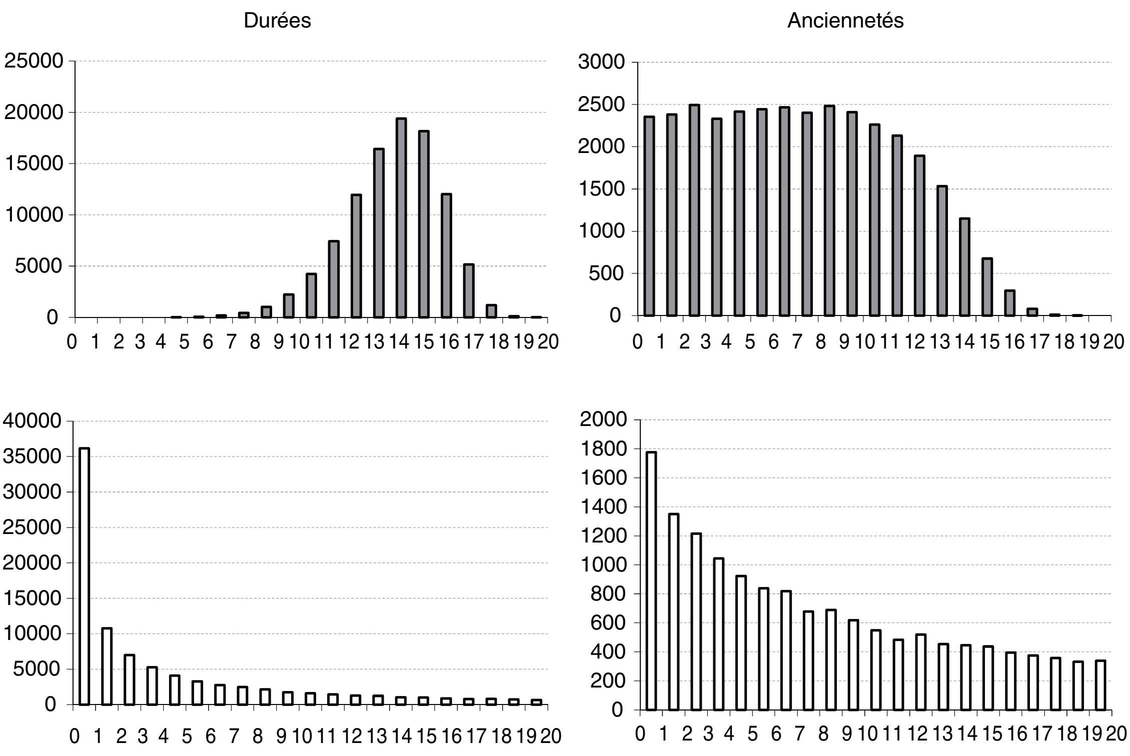
Dans le cas contraire, si le paramètre de forme est supérieur à 1, il y a peu de variabilité des durées : la plupart des personnes ont la même durée à quelques variations près. Celles qui sont en situation monoparentale au moment de l'enquête sont donc à peu près représentatives de l'ensemble des personnes (car il y a très peu d'hétérogénéité en termes de durées). Les anciennetés étant toujours plus courtes que les durées, il en résulte que globalement, les anciennetés des personnes en situation monoparentale au moment de l'enquête seront plus courtes que l'ensemble des durées.

Si l'effet de sélection est plus fort que l'effet de censure, les anciennetés observées seront en moyenne plus grandes que les durées, tandis que si l'effet de censure est plus fort, les anciennetés seront en moyenne plus petites que les durées. Si donc en comparant deux groupes, on trouve qu'en moyenne les anciennetés du premier groupe sont plus faibles que celles du deuxième, cela n'implique pas que les durées sous-jacentes du premier groupe sont en moyenne plus petites que les durées du deuxième groupe. L'ordre des durées ne respectera pas l'ordre des anciennetés si dans le premier groupe l'effet de censure est très fort et dans le deuxième l'effet de sélection est très fort, comme l'illustre la figure II.

L'influence des flux d'entrée sur les anciennetés

La distribution des durées dans la population a donc un impact direct sur les anciennetés. Un autre facteur influence également la distribution des anciennetés observées à un instant donné : il s'agit des flux d'entrée dans la situation de monoparentalité. Si, par exemple, de plus en plus de familles monoparentales se constituent chaque année (flux croissant), les anciennetés observées auront mécaniquement tendance à être courtes, la plupart des familles monoparentales observées dans l'enquête auront été formées peu de temps avant l'enquête. Un flux d'entrée croissant joue donc en faveur du biais de censure (un flux décroissant augmente le biais de sélection). Les simulations présentées dans la figure III illustrent bien ce phénomène.

Figure II
Comparaison de la distribution des durées et des anciennetés entre deux groupes



Note : les durées du groupe 1 (gris foncé), T_1, ont été simulées selon une loi de Weibull(8 , 15) et celles du groupe 2 (blanc) selon une loi de Weibull(0.5 , 5). La distribution des durées est représentée à gauche et celle des anciennetés à droite. Cette figure illustre le fait que les durées peuvent être en moyenne plus grandes dans le groupe 1 que dans le groupe 2 (14 ans contre 10 ans), alors que les anciennetés sont en moyenne plus petites dans le groupe 1 (7 ans) que dans le groupe 2 (12 ans).

Pour résumer, la distribution des anciennetés observées au moment de l'enquête dépend à la fois des flux d'entrée et de la distribution des durées réelles. On a donc l'intuition que si la distribution des anciennetés et les flux d'entrée sont connus, on peut en déduire la distribution des durées. C'est l'objet de la section suivante.

Inférer les durées à partir des anciennetés et des flux d'entrée

Relation entre les flux, les durées et les stocks

On considère dans un premier temps la situation hypothétique dans laquelle les flux d'entrée en monoparentalité sont constants au cours du temps, et où la durée moyenne de la monoparentalité n'évolue pas au cours du temps non plus. Alors les flux de sortie compensent exactement les flux d'entrée et le stock n'évolue pas. Il y a dans ce cas précis, une relation simple entre le stock, les flux et les durées :

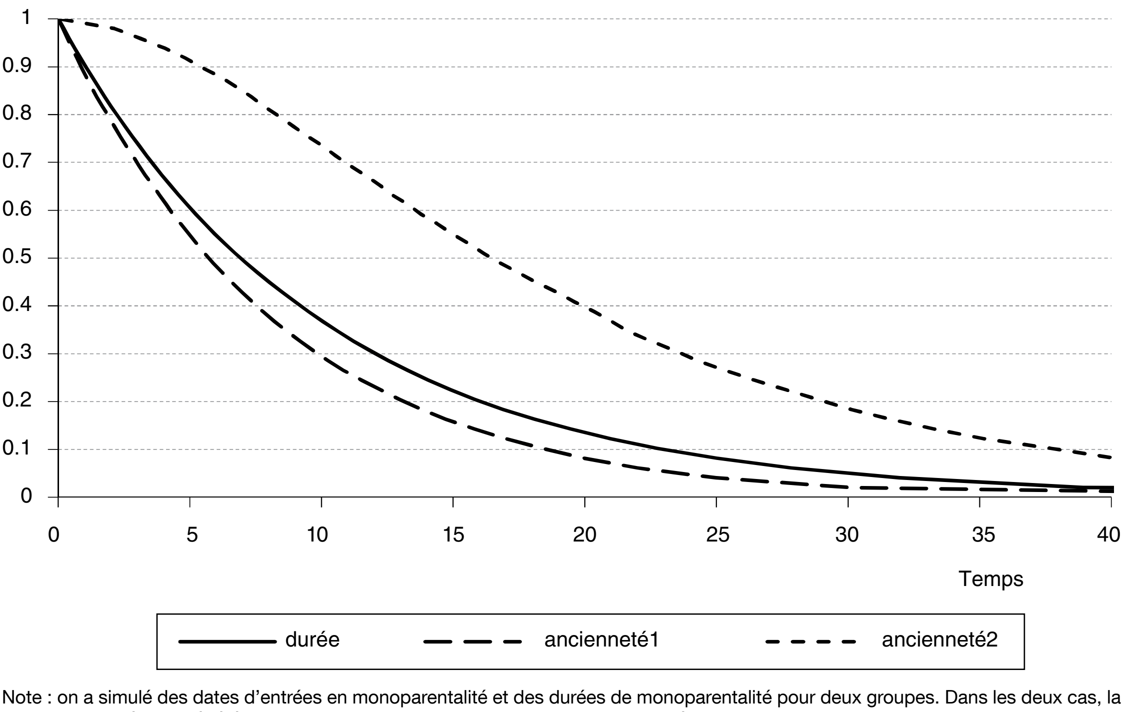
*Stock = Flux * E[T]*, où *E[T]* est l'espérance de la variable aléatoire de durée.

Cette relation se comprend facilement : plus les flux sont importants, plus les stocks observés à un instant donné le sont également ; plus les personnes restent longtemps en situation de monoparentalité, plus elles ont de chance de faire partie du stock, et plus le stock est important. Connaissant les flux et le stock, cette relation nous donne donc directement accès à la durée moyenne passée en situation monoparentale *E[T]*. Dans le cas du régime stationnaire (flux constants) on peut donc directement déduire, à partir des flux d'entrée et de la taille du stock de familles monoparentales, la durée moyenne de la monoparentalité.

En dehors du régime stationnaire, l'égalité n'est plus vérifiée. En fait, si les flux augmentent régulièrement le stock une année donnée est inférieur au dernier flux multiplié par la durée moyenne : Stock < Dernier Flux * E[T], ce qui donne une borne inférieure pour la moyenne des durées. On a la relation inverse dans le cas d'un flux décroissant.

La relation entre les flux, les durées et les stocks ne nous permet toutefois pas de déduire la distribution des durées. Pour cela, il faut en plus connaître la distribution des anciennetés. Comme nous souhaitons connaître la distribution des durées (et pas seulement la durée

Figure III
Influence des flux d'entrées sur les anciennetés observées



Note : on a simulé des dates d'entrées en monoparentalité et des durées de monoparentalité pour deux groupes. Dans les deux cas, la variable de durée est générée selon une loi exponentielle de moyenne 10. La différence entre les deux groupes vient de la distribution des dates d'entrées. Dans le premier groupe, le flux d'entrées en famille monoparentale est croissant tandis que dans le deuxième groupe il est décroissant. Il en résulte que les anciennetés observées A_1 et A_2 ont une distribution différente.

moyenne) et parce que nous ne sommes pas, avec la monoparentalité, dans une situation où les flux sont constants, nous présentons par la suite une méthode plus sophistiquée tenant compte de toute l'information disponible afin de déterminer la distribution des durées quels que soient les flux d'entrée.

Échantillonnage dans le flux et échantillonnage dans le stock

Les modèles de durée classiques, comme le modèle de Cox (1972), se servent la plupart du temps d'échantillonnages dans le flux : on observe les personnes qui entrent dans la situation d'intérêt à une date donnée et on suit ces personnes au cours du temps jusqu'à une date finale. Cela nécessite un dispositif particulier de suivi des personnes et c'est pourquoi ce genre d'études est en général mené sur de petits échantillons en sciences sociales qui nécessitent des enquêtes auprès des personnes, bien qu'on puisse appliquer ces modèles sur des données plus volumineuses si elles sont disponibles. Certaines durées sont alors complètement observées et d'autres sont censurées sur la droite, dans le cas des personnes qui ne sont pas sorties de la situation d'intérêt avant la date finale.

Au contraire, l'échantillonnage dans le stock permet d'étudier un échantillon de personnes qui sont dans la situation d'intérêt à un instant donné, quelle que soit leur date d'entrée dans cette situation. Même si ce type d'échantillonnage est plus rare lorsqu'on s'intéresse à des durées, plusieurs auteurs ont néanmoins développé des modèles de durée sur cette base. Wooldridge (2002) évoque un modèle dans lequel les individus échantillonnés dans le stock sont ensuite suivis pendant une certaine période de temps ; ainsi toutes les durées ne sont pas censurées sur la droite. Kiefer (1988) parle de la difficulté à estimer des durées de chômage à partir du *Current Population Survey* aux États-Unis, en raison du fait que les durées sont à la fois censurées à droite et à gauche. C'est bien souvent pour estimer des probabilités de sortie du chômage, au sein de la théorie de la recherche d'emploi (Atkinson et al., 1984) ou bien en épidémiologie (Keiding, 2006), que des modèles, spécifiques à l'échantillonnage dans le stock, ont été développés. Lorsque toutes les données sont censurées sur la droite, il est toujours possible d'estimer des probabilités de sortie de la situation d'intérêt, comme l'a montré Nickell (1979) dans un article auquel on se référera souvent par la suite, pourvu que l'on connaisse les probabilités d'entrée aux dates précédant l'enquête. Pour cela il a développé un modèle entièrement paramétrique dont l'estimation repose sur le principe du maximum de vraisemblance. Lancaster replace le problème de l'échantillonnage dans le stock dans le contexte plus général des processus de renouvellement (Lancaster, 1990).

La méthode utilisée ici reprend en très grande partie la solution proposée par Nickell (1979), mais plusieurs modifications ont néanmoins été apportées afin de mieux répondre aux questions que l'on se pose sur les durées de la monoparentalité. On reprend l'idée des risques instantanés proportionnels du modèle de Cox que l'on introduit dans le calcul de la vraisemblance donné par Nickell, ce qui rend le modèle proposé ici unique à notre connaissance. De plus, on aura besoin pour estimer les paramètres de ce modèle de disposer d'éléments permettant d'avoir une estimation des flux d'entrée dans cette situation, ce qui se révélera essentiel comme on le verra par la suite. On considère ici les données de l'*EFL* (*Enquête famille et logements*) pour le stock et les données d'*Erfi* (*Étude des relations familiales et intergénérationnelles*) pour les flux.

Principe général de la méthode

Dans le cas de l'*EFL*, on ne connaît que l'année de début de la situation de monoparentalité pour les personnes en famille monoparentale au moment de l'enquête. On ne connaît donc les anciennetés qu'à un an près. Cette incertitude va également se traduire par une incertitude sur les durées inférées. En fait, on choisit ici de modéliser les durées de façon discrète (cf. encadré 2) : on considère la variable aléatoire de durée correspondant à un nombre entier d'années (0, 1, 2, etc.). Par exemple, une personne entrée dans la situation en 2010 et sortie en 2012 aura une durée discrète de 2 ans, alors que sa durée réelle peut valoir entre 1 an (entrée fin 2010 et sortie début 2012) et 3 ans (entrée début 2010 et sortie fin 2012).

La méthode d'inférence consiste dans un premier temps à calculer la fonction de vraisemblance des observations (on observe les anciennetés). On montre que cette fonction de vraisemblance ne dépend que des anciennetés, des flux d'entrée et du risque instantané (voir encadré 3). On choisit ici de modéliser le risque instantané par une fonction constante par morceaux. Bien que le risque instantané soit une fonction discrète, ce choix permet de réduire le nombre de paramètres à estimer et donc d'augmenter la précision des estimations. Il s'agit alors de trouver la fonction constante par morceaux qui maximise la vraisemblance. On peut rendre le modèle plus sophistiqué en faisant dépendre le risque instantané de variables individuelles, comme dans un modèle de Cox (cf. encadré 3). L'inférence consiste alors à trouver les paramètres de régression et la fonction constante par morceaux qui maximisent la vraisemblance.

Encadré 3

MÉTHODE D'INFÉRENCE DES DURÉES À PARTIR DES ANCIENNETÉS ET DES FLUX D'ENTRÉES

On considère m individus, nés entre l'année a_0 et l'année a_1, qui connaissent au cours de leur vie un ou plusieurs épisodes de monoparentalité. Pour chaque épisode, on note D l'année de début, F l'année de fin de la situation et $T = F - D$ la durée (mesurée de façon discrète) de la situation. Si l'entrée en monoparentalité et la sortie de la monoparentalité se font de manière uniforme sur l'année, l'espérance de T est alors égale à l'espérance de la véritable durée. D, F et T sont des variables aléatoires discrètes à valeurs entières.

Hypothèse 1 : on suppose que la variable aléatoire de durée suit la même loi quel que soit le rang de l'épisode de monoparentalité (il peut en effet y avoir plusieurs épisodes de monoparentalité au cours d'une vie).

L'objectif est d'estimer la loi de la variable de durée T conditionnellement à des caractéristiques individuelles x. Pour cela on estimera le risque instantané de cette variable, noté $h(u,x) = P(T = u \mid T \geq u, x)$, pour $u \in \mathbb{N}$.

Au début de l'année e, on réalise une enquête et on observe que n de ces individus ($n \leq m$) sont en situation de monoparentalité. On observe de plus leurs dates d'entrées dans la situation qu'on note $d_1,\ldots, d_n$. C'est à partir de ces dates d'entrées qu'on pourra inférer la loi de T. Pour cela, on calcule dans un premier temps la vraisemblance des observations en faisant l'hypothèse suivante :

Hypothèse 2 : les variables aléatoires T et D sont indépendantes. Cette hypothèse s'interprète comme le fait que la loi de durée de la monoparentalité n'évolue pas au cours du temps : toutes les générations font donc face aux même durées. La contribution à la vraisemblance d'un individu i ayant les caractéristiques x_i et étant entré en monoparentalité l'année d_i s'écrit :

$$P(D_i = d_i \,/\, D_i < e \leq D_i + T_i, x_i) = \frac{P(D_i < e \leq D_i + T_i \,/\, D_i = d_i, x_i)\, A(d_i, x_i)}{P(D_i < e \leq D_i + T_i \,/\, x_i)}$$

où $A(d,x) = P(D = d \,/\, x)$ est la loi conditionnelle de D sachant x.

D'après l'hypothèse 2, on a :

$$P(D_i < e \leq D_i + T_i \,/\, D_i = d_i, x_i) = 1_{(d_i<e)} P(T_i \geq e - d_i \,/\, x_i) = 1_{(d_i<e)} S(e - d_i, x_i)$$

et $P(D_i < e \leq D_i + T_i \mid x_i) = \sum_{u<e} S(e-u)A(u, x_i)$.

Or, $S(t,x) = \prod_{u=0}^{t-1} \left(1 - h(u,x)\right)$.

On propose ici de modéliser le risque instantané comme dans le cas du modèle de Cox à risques instantanés proportionnels :

$$h(t,x) = min\left(h_0(t)exp(\beta x),1\right)$$

Prendre le minimum avec 1 permet de s'assurer que le risque instantané prend bien une valeur entre 0 et 1, comme il se doit dans le cas discret. Enfin, on modélise le risque instantané de base h_0 par une fonction constante par morceaux qu'il faudra estimer. On choisit un pas constant de 3 ans : h_0 est donc constante de 0 à 2 ans, de 3 à 5 ans, etc. jusqu'à 18 ans, puis elle est constante au-delà de 19 ans. Ceci revient donc à estimer 7 paramètres (un pour chaque période de temps sur laquelle la fonction est constante). Le choix d'un pas de 3 ans a été fait pour réduire le nombre de paramètres et augmenter la précision des estimations. L'inconvénient est que cela contraint un peu la forme de la fonction h_0.

La vraisemblance du modèle s'écrit finalement :

$$L(d,x) = \prod_{i=1}^{n} \frac{A(d_i, x_i) \prod_{k=0}^{e-d_i-1} \left(1 - h_0(k)\exp(\beta x_i)\right)}{\sum_{u<e} A(u, x_i) \prod_{k=0}^{e-u-1} \left(1 - h_0(k)\exp(\beta x_i)\right)}$$

où $d=(d_1,\ldots, d_n)$ et $x=(x_1, \ldots, x_n)$.

On cherche alors la fonction constante par morceaux h_0 et le paramètre β qui maximise cette fonction de vraisemblance. On remarque qu'il suffit pour cela de connaître, x étant fixé, la fonction $A(d,x)$ à une constante multiplicative près, ce qui revient à connaître les flux d'entrées en monoparentalité chaque année pour les personnes nées entre l'année a_0 et l'année a_1 et ayant les caractéristiques x. Contrairement au modèle de Cox, on est ici obligé d'estimer conjointement le risque instantané de base h_0 et les paramètres β, bien qu'on ne précise pas de forme paramétrique pour h_0.

D'autres modélisations seraient possibles afin d'établir un lien entre des covariables et le risque instantané. Par exemple, le modèle du temps de défaillance accéléré (*accelerated failure time*) est une alternative intéressante au modèle de Cox. Dans ce modèle, les risques instantanés ne sont plus proportionnels entre eux. L'idée est que chaque personne fait face au même risque de base, mais que le temps ne s'écoule pas à la même vitesse pour chacun : il est soit accéléré, soit ralenti selon que le coefficient associé est plus grand ou plus petit que un. Au final, cela a un effet multiplicatif sur l'espérance de la variable de durée. L'avantage de ce modèle est donc son interprétation très intuitive. L'inconvénient est qu'il faut spécifier de manière paramétrique le risque instantané de base. Or c'est justement ce que l'on évite de faire en ne modélisant pas le risque instantané de base par une fonction déterminée et en laissant les données « librement » estimer ce risque sans contraintes particulières. Tout autre modèle entièrement paramétrique, comme la modélisation *logit* de Nickell, conviendrait également pour modéliser le risque instantané en fonction de covariables. Ainsi, si on a une idée plus ou moins précise de la façon dont le risque instantané évolue au cours du temps et de la façon dont il est modifié selon certaines covariables, il est plus judicieux de construire un modèle entièrement paramétrique qui reflète ces connaissances *a priori*. Si nous avons ici retenu le modèle de Cox, c'est parce que, d'une part, il n'impose pas de donner une forme *a priori* au risque instantané de base et que, d'autre part, l'interprétation des coefficients estimés est simple.

Trois hypothèses ont été faites dans le cadre de cette modélisation. La première est que la durée passée dans la situation de monoparentalité ne dépend pas du rang de l'épisode de monoparentalité au cours de la vie, c'est-à-dire en supposant que, à caractéristiques égales, le fait de vivre un premier ou un deuxième épisode monoparental n'influence pas la durée de la monoparentalité. Il peut en effet arriver de vivre une deuxième fois cette situation, par exemple à la suite d'une remise en couple mettant fin à un épisode de monoparentalité, elle-même suivie d'une séparation ramenant ainsi à la situation de parent isolé. La seconde hypothèse stipule que la durée de la situation est indépendante de la date d'entrée en monoparentalité, ce qui revient à dire qu'il n'y a pas d'effet de génération. Cette hypothèse est forte et n'est sans doute pas vérifiée dans la réalité. Par exemple, le temps pour reformer une union après une séparation se réduit au fil des générations (Costemalle, 2015). Néanmoins nous n'avons pas pu créer un modèle permettant d'estimer de façon robuste cette dépendance. Enfin, la troisième hypothèse est la même que dans un modèle de Cox, à savoir que les risques instantanés sont proportionnels. Cette hypothèse permet de calculer des effets d'une covariable sur le risque instantané indépendamment du temps. On pourra avoir une idée de sa validité en comparant le risque instantané de façon indépendante sur les sous-populations étudiées.

La méthode développée ici est donc originale puisqu'on a combiné deux modélisations connues (celle de Nickell et celle de Cox) pour en faire une troisième. La modélisation de Cox permet d'estimer un risque instantané de base et des paramètres qui indiquent comment évolue le risque instantané par rapport à ce risque instantané de base. Il ne peut malheureusement pas être utilisé dans le cas d'un échantillonnage dans le stock. La modélisation de Nickell, quant à elle, ne permet pas d'estimer un risque instantané de base, ni *a posteriori* une distribution des durées. Son but est de déterminer si une variable externe donnée a une influence positive ou négative sur la probabilité instantanée de sortie d'une situation (initialement chez Nickell, le chômage). Sa force réside dans la simplicité de l'expression du risque instantané (une fonction logistique). L'inconvénient d'une telle modélisation est que l'on force le risque instantané à avoir une certaine forme, ce qui contraint fortement le modèle. Une autre différence entre la modélisation proposée ici et celle de Nickell est que dans le premier cas, les durées suivent des lois discrètes alors que dans le second, elles suivent des lois continues. L'aspect discret du problème ajoute une difficulté supplémentaire.

Résultats de l'inférence sur des simulations

Si on connaît parfaitement les flux d'entrée en monoparentalité, la méthode présentée fonctionne correctement et permet de retrouver la distribution des durées. On peut alors en déduire plusieurs grandeurs associées aux durées comme la moyenne, la médiane ou d'autres quantiles (cf. encadré 2). Les figures IV et V montrent sur des exemples simulés que, quel que soit le biais dominant (biais de censure ou biais de sélection), la méthode d'inférence permet de retrouver les bonnes distributions des durées. Il reste néanmoins trois sources d'incertitude. La première vient du fait qu'on estime des durées discrètes et non continues. La deuxième provient de la taille de l'échantillon : plus celui-ci est petit, plus l'incertitude est grande. Enfin, la troisième concerne le fait qu'on ne connaît pas parfaitement les flux d'entrée mais que ceux-ci sont estimés à partir de l'enquête *Erfi*.

Encadré 4

DÉTERMINATION DES FLUX D'ENTRÉES À PARTIR D'*ERFI*

Pour des caractéristiques individuelles *x* fixées, il s'agit d'estimer *A(d,x)* pour les personnes qui sont nées entre a_0 et a_1, à une constante près (cf. encadré 3). Ceci revient alors à estimer le nombre de personnes ayant les caractéristiques *x* étant entrées dans la situation de monoparentalité l'année *d*. Pour cela, on utilise l'enquête *Erfi* dont les répondants sont nés entre 1926 et 1987. Les répondants à l'*EFL* ont quant à eux pu naître jusqu'en 1992. Il nous manque donc dans *Erfi* le flux d'entrée en monoparentalité des personnes nées entre 1988 et 1992. Pour tenir compte de cela, on estime, dans un premier temps, le flux d'entrée en monoparentalité pour les années inférieures ou égales à 2005 directement à partir de la troisième vague d'*Erfi*. Les personnes nées entre 1988 et 1992 ayant moins de 18 ans sur ces années ne contribuent alors presque pas au flux. Dans un deuxième temps, pour la période 2006-2010 on corrige chaque année le flux par un coefficient multiplicatif afin d'estimer les flux correspondants aux personnes de 18 ans ou plus. Ces coefficients, présentés dans le tableau ci-dessous, sont calculés à l'aide de la répartition des âges d'entrée en monoparentalité (cette répartition est elle-même estimée, à l'aide d'*Erfi*, à partir du flux de personnes entrées en monoparentalité entre 2000 et 2005).

Coefficients correcteurs pour estimer les flux sur la période 2006-2010

	2006	2007	2008	2009	2010
Femmes	1.026	1.026	1.056	1.1	1.125
Hommes	1	1	1	1	1

Source : Ined-Insee, *Erfi*, vague 1, 2005.

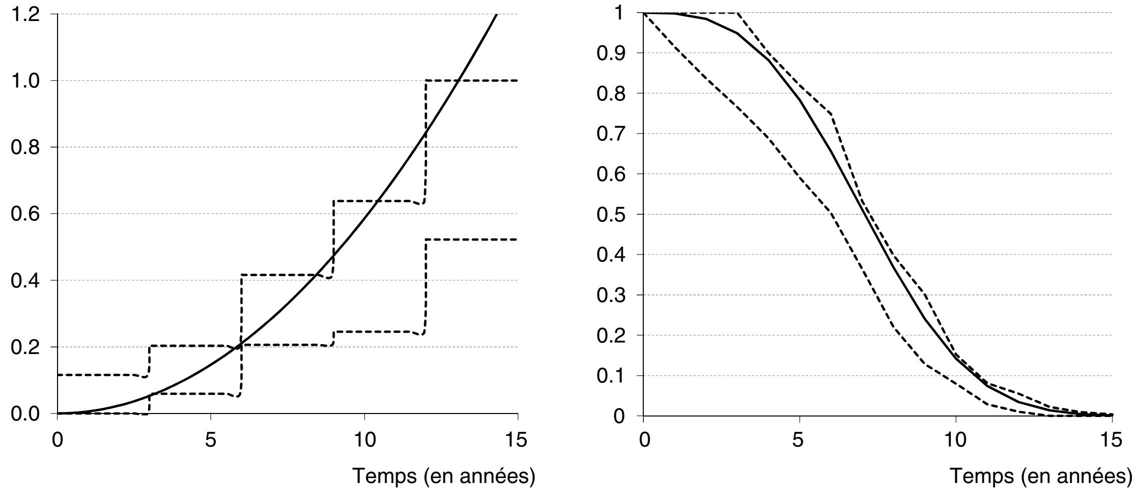
Si une personne connaît plusieurs épisodes de monoparentalité au cours de sa vie, seul le dernier est pris en compte pour la contribution au flux.

Enfin, on lisse les flux ainsi estimés pour atténuer le bruit inhérent aux faibles effectifs présents dans l'enquête *Erfi*. Pour cela on effectue, dans un premier temps, une moyenne mobile sur cinq années qui permet localement d'atténuer les variations aléatoires dues aux petits effectifs. Dans un deuxième temps, on fait une régression polynomiale sur les données ainsi lissées. Les flux d'entrées estimés selon cette méthode sont présentés dans les graphiques en annexe 1.

Influence de la taille de l'échantillon

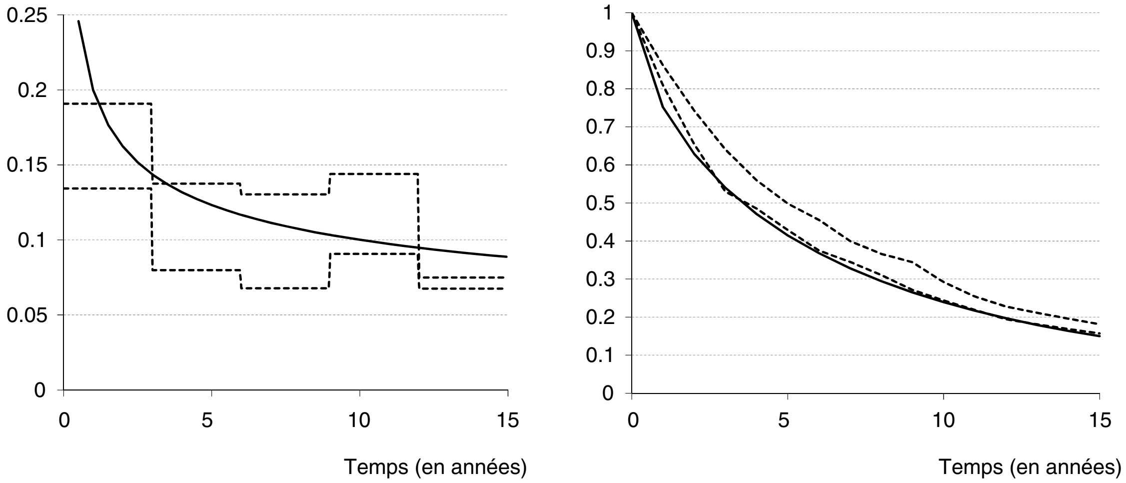
Pour bien comprendre comment évoluent les estimations en fonction de la taille de l'échantillon, on a réalisé une série d'expériences qui consiste pour chacune d'elle à simuler des flux d'entrée et des durées pour un nombre donné d'individus, puis d'en déduire la distribution des durées discrètes (selon la méthode exposée) et enfin d'estimer la moyenne et la médiane des durées. On teste quatre mesures de précision. La première est l'écart entre la moyenne estimée des durées et l'espérance réelle de la variable aléatoire continue de durée ; la deuxième est

Figure IV
Comparaison entre le risque instantané estimé et réel (gauche) et entre la survie estimée et réelle (droite), lorsque le biais de censure est le plus fort



Note : on a généré 100 échantillons de 1 200 observations d'anciennetés à partir d'une loi de Weibull(3 , 8) pour les durées et d'une loi uniforme pour les flux d'entrées. Les courbes en traits pleins représentent le risque instantané et la survie réels. Les courbes en pointillés donnent les intervalles contenant 95 % des estimations.

Figure V
Comparaison entre le risque instantané estimé et réel (gauche) et entre la survie estimée et réelle (droite), lorsque le biais de sélection est le plus fort



Note : on a généré 100 échantillons de 12 300 observations d'anciennetés à partir d'une loi de Weibull(0.7 , 6) pour les durées et d'une loi uniforme pour les flux d'entrées. Les courbes en traits pleins représentent le risque instantané et la survie réels. Les courbes en pointillés donnent les intervalles contenant 95 % des estimations.

l'écart entre la médiane estimée et la médiane réelle ; la troisième est l'écart entre la courbe de survie estimée et la courbe de survie réelle ; la quatrième est l'écart entre le risque instantané estimé et le risque instantané réel. Lorsque la taille de l'échantillon est grande les différentes mesures de précisions ne fluctuent presque pas d'une expérience à l'autre, alors que ce n'est pas le cas lorsque la taille de l'échantillon est petite. Le premier résultat est que la moyenne estimée est inférieure d'environ 0.5 an à la moyenne réelle (quelle que soit la loi de durée sous-jacente). Ainsi, l'inférence des durées conduit à une sous-estimation de la moyenne de 0.5 an, même si la taille de l'échantillon est très importante. Au contraire, la médiane est bien estimée et, lorsque l'échantillon est suffisamment grand, il n'y a presque pas d'écart entre la médiane estimée et la médiane réelle (figure VI). L'écart entre la fonction de survie estimée et la survie réelle et l'écart entre le risque instantané estimé et le risque réel tendent vers 0 lorsque la taille de l'échantillon augmente (figure VII). Ces résultats obtenus à partir de simulations numériques indiquent que l'inférence de la distribution de durées par la méthode de maximisation de la vraisemblance fonctionne correctement et que les résultats deviennent très précis si la taille de l'échantillon est suffisamment importante (de l'ordre de 5 000 personnes).

Par la suite, on ajoutera 0.5 an aux estimations de moyennes afin de tenir compte du biais de l'estimateur de la moyenne. Cette correction s'appuie sur une constatation empirique sur des données simulées, mais n'est pas prouvée théoriquement. On peut néanmoins montrer mathématiquement que si la fonction de survie est bien estimée, comme c'est le cas ici, alors l'erreur sur la moyenne sera d'environ 0.5 an (voir la démonstration en annexe 4).

Estimer les facteurs de risque

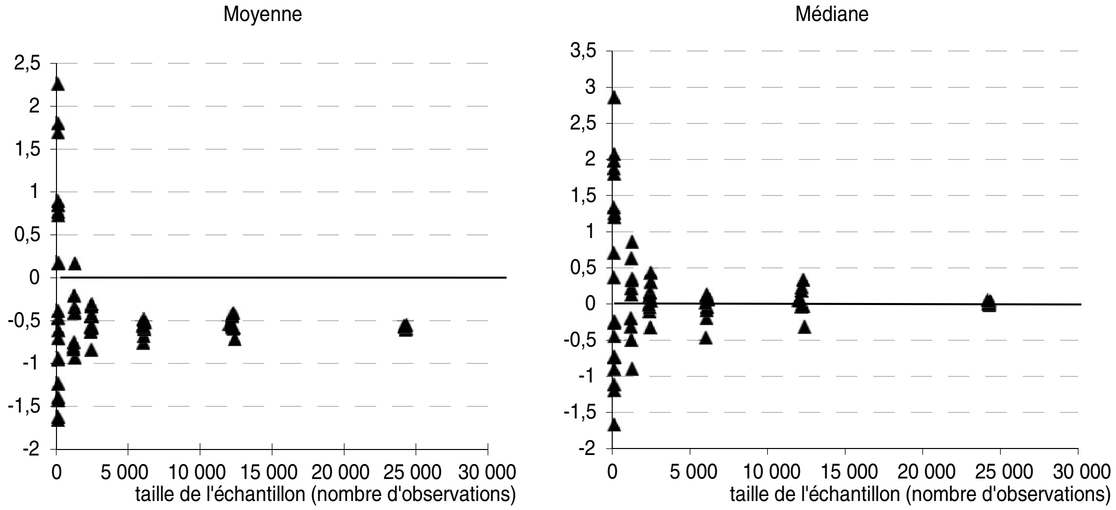
On teste maintenant si le modèle estime correctement les coefficients de régression du risque instantané. On suppose que l'on dispose d'une population formée de trois groupes distincts et faisant chacun face à un risque instantané différent (h_0 pour le premier groupe qu'on désigne comme le groupe de référence, h_1 pour le deuxième et h_2 pour le troisième) :

$$h_1(t) = h_0(t)\exp(\beta_1)\ ;\ h_2(t) = h_0(t)\exp(\beta_2).$$

Ces risques sont, comme l'hypothèse 3 de notre modélisation le suppose, proportionnels entre eux. Il suffit donc d'estimer seulement β_1 et β_2 pour connaître les différences de risques instantanés (et donc de durée) entre les groupes. Si le coefficient est positif, cela signifie que le risque instantané est plus important et que la durée moyenne est donc plus faible. C'est l'inverse si le coefficient est négatif.

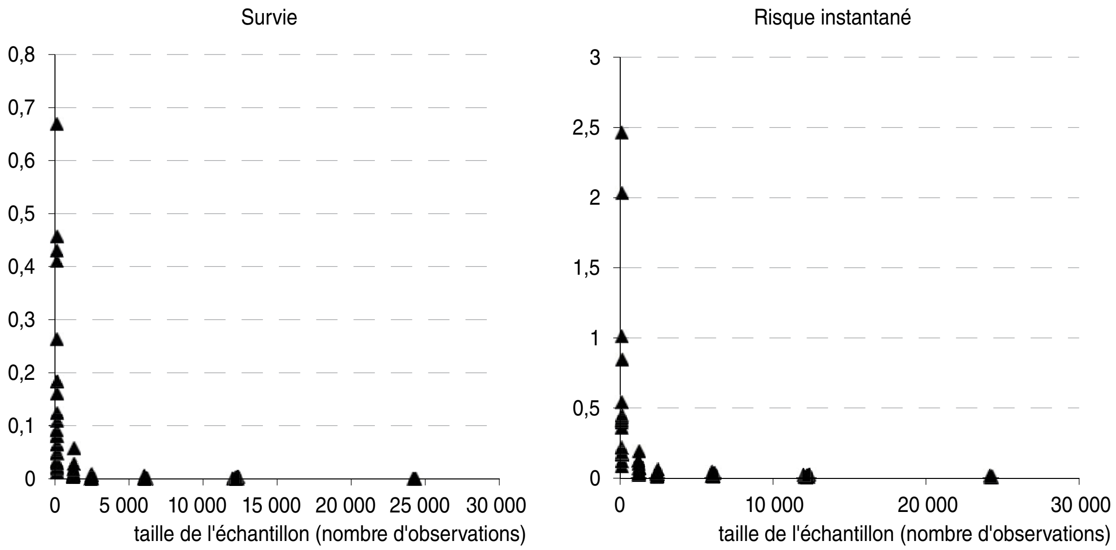
Afin vérifier si notre méthode d'estimation permet de retrouver ces deux paramètres, on simule 200 fois une population totale de 50 000 individus équitablement répartis entre

Figure VI
Écarts entre la moyenne estimée et réelle et entre la médiane estimée et réelle, selon la taille de l'échantillon



Note : chaque point correspond à une expérience, qui consiste à générer des durées de façon aléatoire, à estimer le risque instantané selon la méthode présentée et à en déduire la moyenne ou la médiane. La variable de flux a été générée selon une loi uniforme et la variable de durée selon une loi de Weibull(5 , 7). Ce qui change d'une expérience à l'autre, c'est la taille de l'échantillon qui correspond aux nombre de personnes en famille monoparentale au moment de l'enquête.

Figure VII
Écarts entre la survie estimée et réelle et entre le risque instantané estimé et réel, selon la taille de l'échantillon



Note : chaque point correspond à une expérience. La variable de flux a été générée selon une loi uniforme et la variable de durée selon une loi de Weibull(5 , 7). Si f est la fonction estimée (survie ou risque instantané) et si g est la fonction réelle, on définit l'écart entre ces deux fonctions comme $\sum_{t \geq 0} \left(f(t) - g(t)\right)^2$.

les trois groupes, avec à chaque fois $\beta_1 = 0.5$ et $\beta_2 = -0.5$. Pour chacune de ces 200 simulations, on estime β_1 et β_2 par la méthode du maximum de vraisemblance. On fait ensuite la moyenne de ces estimations sur les 200 expériences. On peut également calculer le taux de couverture, c'est-à-dire la proportion d'estimations telles que la vraie valeur du coefficient se trouve dans l'intervalle de confiance estimé (intervalle de confiance à 95 % de l'estimateur de maximum de vraisemblance, basée sur le comportement asymptotique de l'estimateur qui théoriquement doit suivre une loi normale).

Les résultats de ces simulations sont dans le tableau 1. Elles montrent que le modèle retrouve bien les valeurs des coefficients β_1 et β_2, même si les estimations semblent un peu sous-estimer les

Tableau 1
Moyenne et taux de couverture des estimations des coefficients β_1 et β_2

		β_1	β_2
Vraie valeur		0.5	- 0.5
Estimation	Moyenne	0.47	- 0.48
	Taux de couverture	82.5	85.0

vraies valeurs. Les taux de couverture sont inférieurs à 95 % ce qui indique que les intervalles de confiances estimés sont légèrement trop étroits. Ces taux de couverture ont néanmoins une valeur assez élevée, de l'ordre de 85 %.

Une estimation des durées de la monoparentalité en France

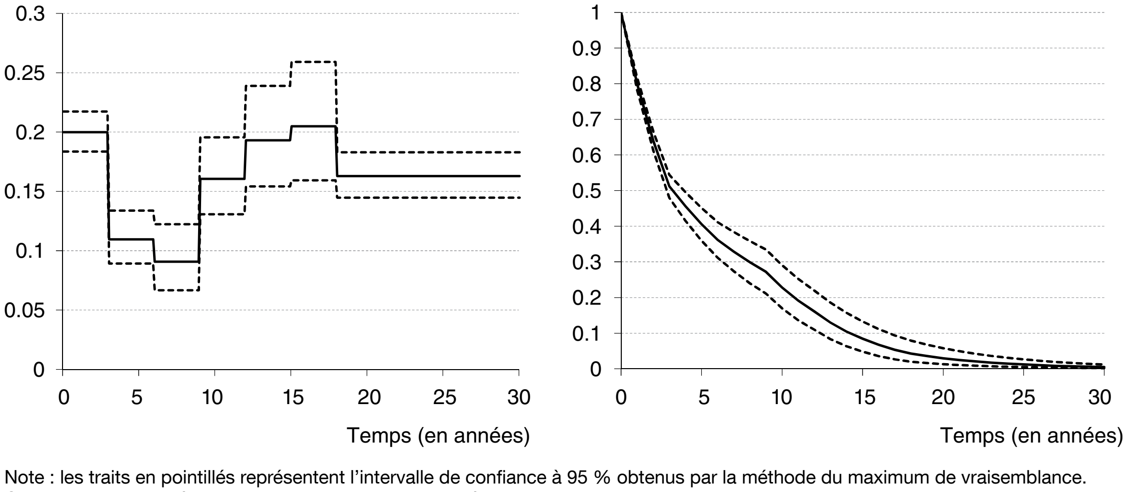
Parmi les 359 770 répondants à l'*EFL*, 12 519 sont en situation de monoparentalité au moment de l'enquête[5], dont 1 073 hommes et 11 446 femmes. Au vu des résultats précédents, on dispose donc d'un échantillon suffisamment grand pour pouvoir inférer la distribution des durées pour les femmes, mais l'échantillon des hommes semble trop petit pour avoir des résultats robustes.

5. *On a retiré les 205 personnes entrées en monoparentalité l'année de l'enquête (en 2011), car elles n'apportent pas d'information à notre modèle, ainsi que les parents vivant en couple non-cohabitant, situation dont on ne peut déterminer l'ancienneté.*

Les durées estimées : une forme « en U »

On présente ici les résultats des estimations du risque instantané de sortir de la monoparentalité sans prendre en compte de covariables, c'est-à-dire en ne tenant compte que du temps, sans introduire d'autres facteurs qui pourraient influencer la probabilité de sortir de la monoparentalité. La figure VIII montre que le risque instantané global n'est pas monotone : initialement il diminue, puis se stabilise avant d'augmenter. Cette forme en « U » suggère que soit on sort rapidement de la monoparentalité, soit on y reste longtemps. En effet, la probabilité de sortir de cette situation est la plus faible entre 3 et 8 ans. La courbe de survie obtenue à partir du risque instantané montre qu'au bout de 3 ans la moitié des parents de famille monoparentale sont sortis de cette situation ; au bout de 8 ans il en reste encore 30 %, au bout de 12 ans 16 %, et seulement 4 % y restent plus de 18 ans. On peut aussi déterminer la moyenne de la durée passée dans cette situation (cf. encadré 2) : on trouve ici que la durée moyenne est de 5.7 ans. La moyenne des anciennetés quant à elle est de

Figure VIII
Estimation du risque instantané et de la survie associés aux durées de monoparentalité

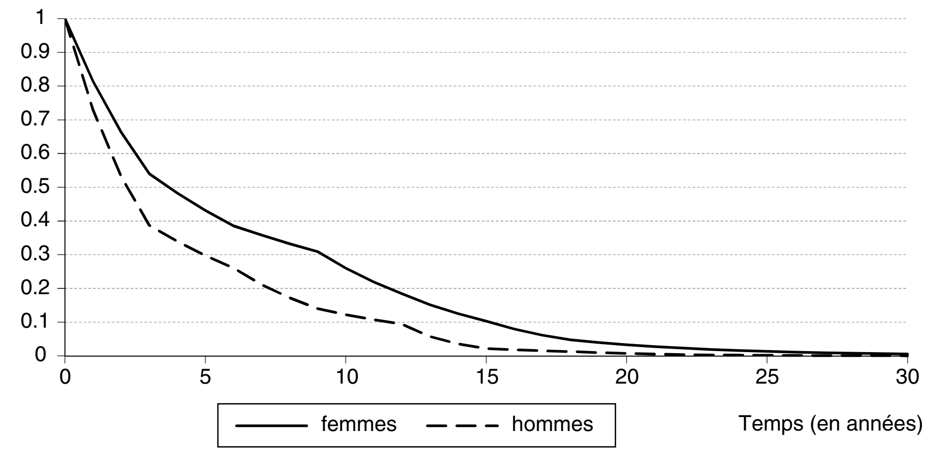


Note : les traits en pointillés représentent l'intervalle de confiance à 95 % obtenus par la méthode du maximum de vraisemblance.
Champ : parents isolés avec enfants mineurs, France métropolitaine.
Source : Insee, *Enquête famille et logements* (*EFL*) 2011 et Ined-Insee, *Étude des relations familiales et intergénérationnelles* (*Erfi*), vagues 1 et 3, 2005 et 2011.

Figure IX
Risque instantané de sortie de la monoparentalité des hommes et des femmes

Champ : parents isolés avec enfants mineurs, France métropolitaine.
Source : Insee, *Enquête famille et logements* (*EFL*) 2011 et Ined-Insee, *Étude des relations familiales et intergénérationnelles* (*Erfi*), vagues 1 et 3, 2005 et 2011.

Figure X
Fonctions de survie associées aux durées passées en famille monoparentale, pour les hommes et les femmes



Champ : Parents isolés avec enfants mineurs, France métropolitaine.
Source : Insee, *Enquête famille et logements* (*EFL*) 2011 et Ined-Insee, *Étude des relations familiales et intergénérationnelles* (*Erfi*), vagues 1 et 3, 2005 et 2011.

Tableau 2
Estimations de la survie, de la moyenne et de la médiane des durées passées en famille monoparentale

	Moyennes			Médianes		
	Valeur estimée	Intervalles de confiance		Valeur estimée	Intervalles de confiance	
		Maximum de vraisemblance	*Bootstrap*		Maximum de vraisemblance	*Bootstrap*
Ensemble	5.7	[4.9 ; 6.6]	[5.5 ; 6.1]	3.2	[2.8 ; 3.9]	[2.8 ; 3.9]
Hommes	4.1	[2.8 ; 6.0]	[3.7 ; 4.7]	2.2	[1.8 ; 2.8]	[1.8 ; 3.0]
Femmes	6.1	[5.2 ; 7.2]	[5.9 ; 6.5]	3.7	[3.1 ; 4.6]	[3.0 ; 4.3]

Note : la moyenne est estimée ici selon l'estimateur présentée dans l'encadré 2 corrigée du biais de 0.5 an. Les intervalles de confiance obtenus par *bootstrap* reposent sur 320 échantillons tirés aléatoirement selon la distribution des anciennetés observées (tirage aléatoire avec remise). Pour chacun de ces échantillons on a estimé le risque instantané dont on a déduit la moyenne et la médiane des durées.
Champ : France métropolitaine.
Source : Insee, *Enquête famille et logements* (*EFL*) 2011 et Ined-Insee, *Étude des relations familiales et intergénérationnelles* (*Erfi*), vagues 1 et 3, 2005 et 2011.

5.5 ans. Elle est donc proche de la moyenne des durées estimées. On est ainsi dans une situation particulière où le biais de censure et le biais de sélection se neutralisent presque.

En estimant séparément le risque instantané pour les femmes et les hommes, on constate qu'ils n'ont pas la même forme ; par conséquent, l'hypothèse de proportionnalité des risques instantanés n'est pas appropriée pour la comparaison entre sexes (figures IX et X). Si le risque instantané des femmes a la forme en « U » observée pour l'ensemble, celui des hommes est plus fluctuant et plus élevé en moyenne.

Cette fluctuation laisse penser qu'il n'y a pas assez d'observations pour les hommes dans l'*EFL* : autour de 1 000. D'autre part, le fait que dans l'ensemble, il y a beaucoup plus de femmes que d'hommes (à peu près dix fois plus) explique de plus qu'on retrouve la forme en « U » à la fois pour l'ensemble des parents en famille monoparentale et pour les femmes seulement.

La durée moyenne de la monoparentalité estimée ici est de 6.1 ans pour les femmes et 4.1 ans pour les hommes (tableau 2). Des intervalles de confiance à 95 % ont été estimés successivement par la méthode du maximum de vraisemblance, puis par *bootstrap*. En ce qui concerne la moyenne, ces derniers sont plus étroits que ceux donnés par la méthode du maximum de vraisemblance. En revanche, pour la médiane, on obtient presque les mêmes intervalles de confiance. Ceci constitue une indication supplémentaire de la fiabilité des estimations.

Déterminer la durée moyenne à partir des flux et des stocks

On a pu déterminer des moyennes de durées de monoparentalité à partir de la distribution des durées estimée. Comme on l'a expliqué auparavant, il est néanmoins possible d'avoir une estimation de la borne inférieure de la moyenne des durées à l'aide de la relation entre les stocks, les flux et les durées.

On peut appliquer ce principe à l'*EFL* car on connaît la taille du stock de familles monoparentales en 2011 (1 449 000) ainsi que la grandeur du dernier flux (254 000 entrées en 2010). On en déduit alors que la durée moyenne de la monoparentalité doit être supérieure à 6 ans pour les femmes et 4.4 ans pour les hommes (tableau 3), or nos estimations précédentes sont de 6.1 ans pour les femmes et 4.1 ans pour les hommes.

Le dernier flux est sans doute sous-estimé car on ne prend pas en compte les personnes qui sont entrées dans la situation en 2010 et qui en sont sorties la même année. Il en résulte que l'estimation de la borne inférieure de la moyenne de la durée passée en situation monoparentale est légèrement sous-estimée également. Au final, ces différentes estimations semblent tout à fait cohérentes entre elles.

Les durées et les anciennetés correspondent à deux concepts différents. Mais dans le cas présent, la répartition des anciennetés est assez proche de celles des durées. On est en effet dans une situation où le biais de censure et le biais de sélection se compensent approximativement, les anciennetés peuvent donc fournir une première approximation des durées.

Des différences sensibles des durées estimées selon la cause d'entrée dans la monoparentalité

On s'intéresse maintenant aux durées obtenues en introduisant des covariables indépendantes du temps. Compte tenu de la faible taille de l'échantillon d'hommes, on se restreint seulement aux résultats concernant les femmes. Les durées de monoparentalité sont estimées successivement selon la cause d'entrée dans la situation, selon le niveau de diplôme atteint

Tableau 3
Stock et flux de familles monoparentales

	Stock en 2011	Flux en 2010	Stock/Flux
Hommes	208 904	47 977	4.4
Femmes	1 239 843	206 067	6.0

Champ : France métropolitaine.
Source : Insee, *Enquête famille et logements* 2011.

au moment de l'enquête et selon la catégorie sociale au moment de l'enquête. À chaque fois, on présente les estimations des coefficients β du modèle à risques instantanés proportionnels, par rapport à un groupe de référence ; si β est positif, cela signifie que le risque instantané de sortie de la situation monoparentale est plus élevé que celui du groupe de référence, donc que sa survie dans la situation (la durée) est plus courte que celle du groupe de référence. Les durées moyennes sont déduites des risques instantanés estimés pour chaque groupe, eux-mêmes déduits du risque instantané de base (voir annexe 2) et des risques relatifs.

Les causes d'entrée en monoparentalité sont au nombre de trois : la séparation d'avec le conjoint (78 %) un enfant né hors de la vie en couple (16 %), le décès du conjoint (6 %). Des différences de durée notables sont observées entre ces causes d'entrée. Par rapport à une entrée associée à un enfant né hors de la vie de couple, la probabilité instantanée de sortie de la monoparentalité est 1.8 fois plus élevée lorsque la cause d'entrée est une séparation, et 1.7 fois plus élevée lorsque que c'est à la suite du décès du conjoint (tableau 4.A). Les femmes qui ont vécu une situation de monoparentalité à l'issue d'une séparation sont donc celles qui y passent le moins de temps (5.4 ans), suivies des femmes qui se trouvent à la tête d'une famille monoparentale suite à un veuvage (5.7 ans) et celles qui ont eu un enfant hors couple restent nettement plus longtemps dans cette situation (9.1 ans).

Des différences apparaissent également selon le niveau de diplôme atteint. On distingue les personnes n'ayant aucun diplôme, les personnes ayant un diplôme de niveau inférieur au baccalauréat, celles ayant un diplôme de niveau équivalent au baccalauréat et celles ayant un diplôme de niveau strictement supérieur au baccalauréat. La probabilité instantanée de sortie de la monoparentalité apparaît toujours plus élevée pour les femmes ayant un diplôme, ce qui signifie que celles qui n'en ont aucun restent en moyenne plus longtemps en situation monoparentale que les autres (tableau 4.B). La probabilité de sortie ne s'élève cependant pas régulièrement selon le niveau de diplôme atteint ; ainsi, d'après ces estimations, ce sont les femmes qui ont un diplôme inférieur au baccalauréat et celles qui ont seulement le baccalauréat pour lesquelles les durées sont les moins

Tableau 4
Estimation des risques relatifs de sortie de la monoparentalité et des durées moyennes dans la situation

A. Selon la cause d'entrée

Cause d'entrée	Risque relatif	valeur-p	Durée moyenne
Enfant hors couple (16 %)	1	-	9.1
Séparation (78 %)	1.84	7.90E-77	5.4
Veuvage (6 %)	1.73	6.00E-31	5.7

B. Selon le niveau de diplôme

Diplôme	Risque relatif	valeur-p	Durée moyenne
Aucun diplôme (21 %)	1	-	7.6
Inférieur au bac (34 %)	1.40	2.10E-28	5.5
Baccalauréat (19 %)	1.38	6.70E-18	5.6
Supérieur au baccalauréat (27%)	1.20	4.70E-07	6.4

C. Selon la catégorie sociale

Catégorie sociale	Risque relatif	valeur-p	Durée moyenne
Ouvriers (11 %)	1	-	6.9
Artisans, commerçants, chefs d'entreprises (3 %)	1.02	7.70E-01	6.8
Cadres et profession intellectuelles supérieures (9 %)	0.90	1.00E-01	7.5
Professions intermédiaires (23%)	1.05	2.90E-01	6.6
Employés (54%)	1.11	7.60E-03	6.3

Note : les pourcentages entre parenthèses donnent la proportion de chaque catégorie.
Champ : femmes, France métropolitaine.
Source : Insee, *Enquête famille et logements* (*EFL*) 2011 ; Ined-Insee, *Étude des relations familiales et intergénérationnelles* (*Erfi*), vagues 1 et 3, 2005 et 2011.

élevées (respectivement 5.5 et 5.6 ans) contre 7.6 ans pour celles n'ayant aucun diplôme et 6.4 ans pour celles qui ont un diplôme supérieur au baccalauréat.

On distingue enfin cinq catégories sociales[6] : les artisans, commerçants ou chefs d'entreprise, les cadres, les professions intermédiaires, les employés et les ouvriers. La catégorie sociale semble avoir très peu d'effet sur la durée passée en situation de monoparentalité (tableau 4.C). Le seul écart de probabilité statistiquement significatif (au seuil de 5 %) s'observe pour les femmes employées, avec une probabilité instantanée de sortie de la monoparentalité légèrement plus élevée que les ouvrières. En moyenne, les femmes artisanes, commerçantes ou chefs d'entreprises restent 6.8 ans en famille monoparentale, les cadres 7.5 ans, les professions intermédiaires 6.6 ans, les employées 6.3 ans et les ouvrières 6.9 ans.

Les différences de durée les plus importantes sont ainsi observées entre les causes d'entrée dans la monoparentalité. Ces différences de durée selon la cause d'entrée sont associées en premier lieu à l'âge des enfants au début de la situation ; en effet, la durée maximale de la monoparentalité est limitée par l'âge du plus jeune des enfants au moment où la famille devient monoparentale. Les femmes qui ont eu un enfant sans avoir vécu en couple sont en situation monoparentale d'emblée, dès la naissance de l'enfant. Si elles ne forment pas de couple par la suite, elles peuvent ainsi rester dans cette situation 18 ans, voire plus longtemps, dans le cas (rare) où elles auraient plusieurs enfants sans jamais vivre en couple. Les femmes qui deviennent mères de famille monoparentale à la suite d'une séparation ou d'un décès ont, à l'entrée dans cette situation, des enfants plus âgés.

La cause d'entrée pourrait contribuer par ailleurs aux écarts de durée estimés selon les niveaux de diplôme : en effet, pour les femmes observées en situation de monoparentalité en 2011, l'entrée dans la situation est associée au fait d'avoir eu un enfant hors couple plus souvent lorsqu'elles n'ont pas de diplôme que lorsqu'elles ont un diplôme supérieur au baccalauréat (respectivement 20 % et 11 % des cas).

6. *Les retraités et les chômeurs ayant déjà travaillé sont reclassés dans leur ancienne catégorie sociale. Les chômeurs n'ayant jamais travaillé, les inactifs de moins de 60 ans, les militaires, les étudiants les personnes sans activité de plus de 60 ans ainsi que les agriculteurs ne sont pas pris en compte.*

Si l'âge des enfants à l'entrée dans la monoparentalité influence mécaniquement la durée dans la situation, un autre facteur peut aussi contribuer à cette durée : la mise ou remise en couple du parent, qui interrompt instantanément la monoparentalité. On sait par exemple qu'après une séparation, les femmes plus diplômées ne reforment pas plus rapidement un couple que les moins diplômées (Costemalle, 2015). Mais l'influence des mises ou remises en couple sur les écarts de durée de monoparentalité selon la cause d'entrée ne peut pas être évaluée ici car les données ne permettent pas ici de connaître la cause de sortie de la monoparentalité.

Restent quelques sources d'incertitude. L'une porte sur les durées de monoparentalité concernant les hommes ; on a vu que les effectifs d'hommes dans cette situation sont trop faibles pour pouvoir mener des inférences précises (les simulations ont montré que c'est à partir d'un échantillon d'environ 5 000 individus que les estimations deviennent précises), les résultats sont donc fragiles. Les autres sources d'incertitude concernent à la fois les estimations pour les hommes et pour les femmes : l'une est que l'on ne connaît les anciennetés qu'à un an près. D'après les simulations présentées, le fait de ne connaître les anciennetés qu'à un an près semble résulter en un biais de 0.5 an environ pour l'estimation de la moyenne des durées, mais cela n'affecte pas l'estimation de la médiane. Si l'on veut tenir compte de ce résultat observé à partir de plusieurs cas simulés, il faut donc rajouter 0.5 an aux estimations des durées moyennes. La seconde concerne l'estimation des flux d'entrée dans la situation : ils ont été estimés à partir d'une autre source, *Erfi*, dont les effectifs sont beaucoup plus faibles que ceux de l'*EFL* et l'impact d'une erreur d'estimation de ces flux sur les résultats finaux n'est pas mesuré ici.

* *
*

Revenons pour finir sur la méthode proposée pour l'estimation des durées passées en famille monoparentale, et d'abord sur les trois hypothèses formulées dans le cadre de la modélisation.

La première est que la durée de monoparentalité ne dépend pas du rang de l'épisode de monoparentalité. Cette hypothèse semble *a priori* peu vraisemblable, car on peut penser que lors d'une deuxième expérience de monoparentalité, les enfants sont plus âgés, ce qui limite mécaniquement la durée passée en famille monoparentale pour les personnes qui ne reforment pas un couple cohabitant (d'après la définition retenue ici, les enfants doivent être âgés de moins de 18 ans). Néanmoins, la majorité des adultes étant à la tête d'une famille monoparentale ne vivent qu'une fois cette situation. Selon l'enquête *Erfi*, seules 16 % des personnes entre 18 et 72 ans en 2005 ayant connu au moins un épisode de monoparentalité au cours de leur vie, ont vécu deux fois cette situation. L'hypothèse formulée apparait donc raisonnable.

La deuxième hypothèse suppose que la loi de la durée passée en famille monoparentale n'évolue pas avec le temps. Ceci semble peu probable, car les raisons de la formation de familles monoparentales évoluent avec le temps : il y a de moins en moins de veuves et de femmes ayant eu un enfant hors couple et de plus en plus de personnes séparées (selon l'enquête *Erfi*). On a vu que ces différentes causes d'entrée présentent des durées hétérogènes, ce qui fait que les durées passées en situation monoparentale évoluent avec la structure de la population des familles monoparentales, donc changent avec le temps. Il y a alors une corrélation entre la date d'entrée en famille monoparentale et la durée de la monoparentalité. On a tenté de comprendre l'effet de cette corrélation sur les estimations en s'appuyant sur des simulations (voir annexe 3), qui génèrent de façon aléatoire des durées corrélées soit positivement, soit négativement, avec la date d'entrée en famille monoparentale. Ainsi, si la durée moyenne de monoparentalité diminue avec la date d'entrée dans la situation, alors les estimations du modèle sous-estiment la moyenne des durées et au contraire, si la corrélation est positive, on surestime la moyenne.

Enfin la troisième hypothèse porte sur la proportionnalité des risques instantanés. Contrairement au modèle de Cox, nous ne disposons pas ici de test permettant de confirmer ou d'infirmer cette hypothèse. Il est néanmoins possible d'avoir une idée de sa validité en estimant séparément les courbes de risque instantané pour chaque sous-population. On a vu par exemple que cette hypothèse ne semble pas être vérifiée lorsque l'on veut distinguer les femmes des hommes.

Plus généralement, il convient de mentionner certaines limites de la méthode d'estimation développée dans cet article. Premièrement, il faut disposer d'échantillons de grande taille pour pouvoir obtenir des résultats fiables. C'est une limite importante, car il n'est alors pas possible d'appliquer cette méthode sur des enquêtes ou sur des sous-populations trop petites. Au niveau de la modélisation ensuite, il apparaît difficile de prendre en compte conjointement plusieurs variables explicatives, comme c'est normalement le cas dans un modèle de Cox à risques instantanés proportionnels. Il faudrait en effet pour cela pouvoir estimer des flux d'entrée selon le croisement de plusieurs variables, or les effectifs de l'enquête *Erfi* ne permettent pas d'avoir une si grande précision. Il en résulte qu'on n'a pas pris ici en compte l'interaction des différentes variables explicatives sur la probabilité instantanée de sortir de la monoparentalité. Il n'est donc pas possible d'estimer des effets « toutes choses égales par ailleurs ». De plus, même si la modélisation du risque instantané par une fonction constante par morceaux laisse beaucoup de liberté, on a contraint ce risque instantané à être constant sur des périodes de 3 ans, pour les besoins de l'estimation. En effet, s'il y a trop de paramètres à estimer pour le risque instantané, cela peut de nuire à la précision des estimations.

La limite principale de la méthode présentée reste néanmoins qu'il faut connaître les flux d'entrée dans la situation que l'on étudie. Or ces flux ne peuvent pas être déduits d'une enquête par échantillonnage dans le stock. Il faut donc disposer d'une autre source d'information pour établir ces flux. La nécessité de connaître les flux associés à la situation d'intérêt est ainsi le point faible de la démarche. Toutefois, si l'on ne connaît pas ces flux, il est toujours possible soit d'en avoir une idée *a priori*, soit d'élaborer plusieurs scénarios. On pourrait même imaginer un modèle bayésien qui, partant d'une distribution *a priori* sur les flux d'entrée, estimerait une distribution *a posteriori* des durées.

Dans de nombreux domaines, les durées sont très difficiles à mesurer car la plupart des données résultent d'échantillonnages dans le stock. Malgré ces limites, la méthode développée dans cet article reste intéressante, offrant un moyen simple d'obtenir des estimations sur les durées à partir des anciennetés observées. □

BIBLIOGRAPHIE

Algava, E. (2002). Les familles monoparentales en 1999. *Population*, 57(4-5), 733–758.

Algava, E., Le Minez, S., Bressé, S. & Pla, A. (2005). Les familles monoparentales et leurs conditions de vie. Drees, *Études et Résultats* N° 389.
http://drees.social-sante.gouv.fr/etudes-et-statistiques/publications/etudes-et-resultats/article/les-familles-monoparentales-et-leurs-conditions-de-vie.

Atkinson, A. B., Gomulka, J. & Micklewright, J. (1984). Unemployment benefit, duration and incentives in Britain. *Journal of Public Economics*, 23, 3–26.

Boiron, A., Huwer, M. & Labarthe, J. (2016). Inégalités de niveaux de vie et pauvreté en 2013. Insee Références *Les revenus et le patrimoine des ménages édition 2016*, 9–21.
https://www.insee.fr/fr/statistiques/2017606?sommaire=2017614.

Breuil-Genier, P., Buisson, G., Robert-Bobée, I. & Trabut, L. (2016). Enquête Famille et Logements adossée au Recensement de 2011 : comment s'adapter à la nouvelle méthodologie des enquêtes annuelles et quels apports ? *Économie et Statistique*, 483-484-485, 205–226.
https://www.insee.fr/fr/statistiques/2017652?sommaire=2017660.

Buisson, G., Costemalle, V. & Daguet, F. (2015). Depuis combien de temps est-on parent de famille monoparentale ? *Insee Première* N°1539
https://www.insee.fr/fr/statistiques/1283845

Bumpass, L. L. & Raley, R. K. (1995). Redefining single-parent families: cohabitation and changing family reality. *Demography*, 32(1), 97–109.

Chambaz, C. (2000). Les familles monoparentales en Europe : des réalités multiples. Drees, *Études et Résultats* N° 66.
http://drees.social-sante.gouv.fr/IMG/pdf/er066.pdf.

Chardon, O., Daguet, F. & Vivas, E. (2008). Les familles monoparentales – des difficultés à travailler et à se loger. *Insee Première* N° 1195.
https://www.insee.fr/fr/statistiques/1281271.

Costemalle, V. (2015). Parcours conjugaux et familiaux, des hommes et des femmes selon les générations et les milieux sociaux. Insee Références *Couples et familles*, 63–76.
https://www.insee.fr/fr/statistiques/2017510?sommaire=2017528.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, series B*, 34(2), 187–220.

David, O., Eydoux, L., Ouallet, A. & Séchet, R. (2003). Les familles monoparentales : une perspective internationale. L*'e-ssentiel* N° 15.
https://www.caf.fr/sites/default/files/cnaf/Documents/Dser/essentiel/15%20ESSENTIEL%20MONOPARENTALITE.pdf.

David, O., Eydoux, L., Martin, C., Millar, J. & Séchet, R. (2004). Les familles monoparentales en Europe. *Dossiers d'études de la CAF* N° 54.
https://www.caf.fr/sites/default/files/cnaf/Documents/Dser/dossier_etudes/dossier_54_-_familles_monoparentales.pdf.

Eydoux, A. & Letablier, M.-T. (2009). Familles monoparentales et pauvreté en Europe : quelles réponses politiques ? [L'exemple de la France, de la Norvège et du Royaume-Uni]. *Politiques sociales et familiales*, 98, 21−35.
http://www.persee.fr/doc/caf_2101-8081_2009_num_98_1_2487

Haut Conseil de la Famille (2014). Les ruptures familiales − État des lieux et propositions. Rapport du 10 avril 2014.
http://www.ladocumentationfrancaise.fr/rapports-publics/144000594/index.shtml.

Keiding, N. (2006). Event history analysis and the cross-section. *Statistics in Medicine*, 25, 2343–2364.

Kiefer, N. M. (1988). Duration data and hazard functions. *Journal of Economic Literature*, 26(2), 646–679.

Lancaster, T. (1990). *The econometric analysis of transition data*. Cambridge University Press.

Le Pape, M.-C., Lhommeau, B. & Raynaud, E. (2015). Les familles monoparentales en Europe : de nouvelles façons de faire famille pour de

nouvelles normes ? Insee Références *Couples et familles*, 27–40.
https://www.insee.fr/fr/statistiques/2017504?sommaire=2017528.

McKay, S. (2002). The dynamics of lone parents, employment and poverty in Great Britain. *Sociology and Social Politics*, 6-2, 101-124.

Nickell, S. (1979). Estimating the probability of leaving unemployment. *Econometrica*, 47(5), 1249–1266.

Rabier, R. (2014). Les familles monoparentales, souvent en situation de précarité. *Insee Analyses Languedoc-Roussillon* N° 2.
https://www.insee.fr/fr/statistiques/1285832.

Régnier-Loilier, A. (2012). Présentation, questionnaire et documentation de la troisième vague de l'Étude des relations familiales et intergénérationnelles (Érfi-GGS 2011). Ined, Document de travail N° 187.
https://www.ined.fr/fr/publications/document-travail/etude-relations-familiales-intergenerationnelles/.

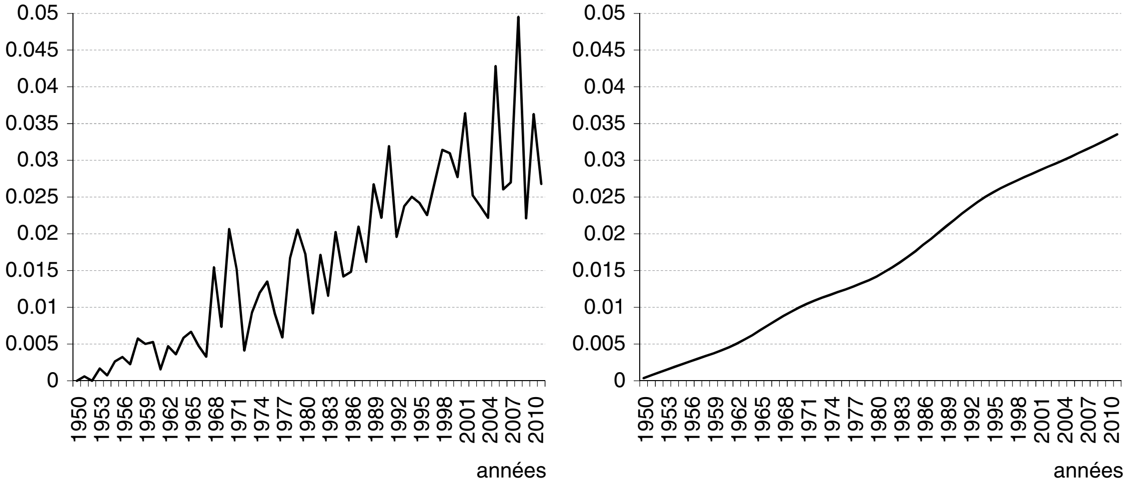
Toulemon, L. (2011). Individus, familles, ménages, logements : les compter, les décrire. *Travail, Genre et Société*, 26, 47–66.

Vespa, J., Lewis, J. M. & Kreider, R. M. (2013). America's families and living arrangements: 2012. *Current Population Reports*, P20-570. Washington, DC: U.S. Census Bureau.
https://www.census.gov/library/publications/2013/demo/p20-570.html.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge (Ma): The MIT Press.

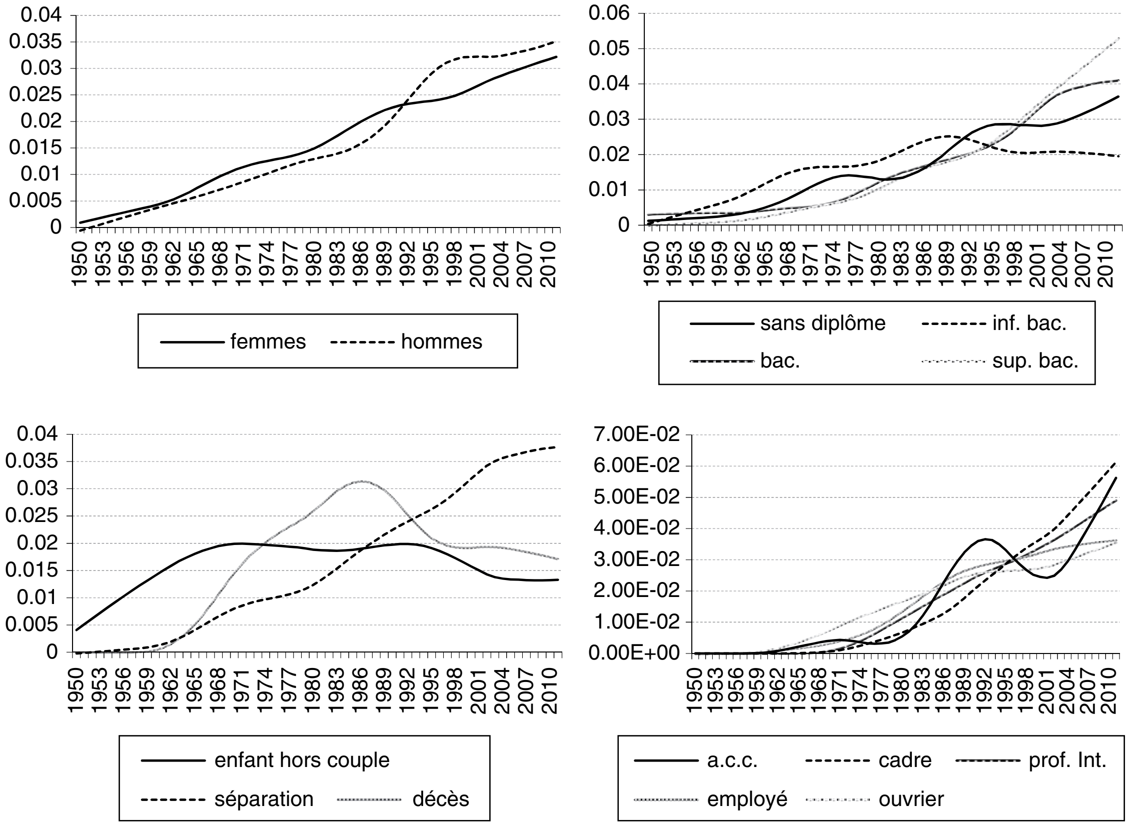
LES ESTIMATIONS LISSÉES DES FLUX D'ENTRÉE

Figure A1-I
Estimations des flux annuels d'entrée en famille monoparentale, avant et après lissage



Note : le lissage est obtenu par une moyenne mobile suivie d'une approximation polynomiale.
Champ : France métropolitaine.
Source : Ined-Insee, *Erfi*, vagues 1 (2005) et 3 (2011).

Figure A1-II
Estimations des flux annuels d'entrée en famille monoparentale, pour quelques caractéristiques individuelles



Note : le lissage est obtenu par une moyenne mobile suivie d'une approximation polynomiale. La catégorie sociale a.c.c. correspond aux artisans, commerçants et chefs d'entreprises.
Champ : France métropolitaine.
Source : Ined-Insee, *Erfi*, vagues 1 (2005) et 3 (2011).

ANNEXE 2

ESTIMATIONS DES FONCTIONS DE SURVIE ASSOCIÉES AUX DURÉES DE MONOPARENTALITÉ

Tableau A2-1
Fonctions de survie estimées dans le modèle sans covariables

	Ensemble			Hommes			Femmes		
	estimation	borne inférieure	borne supérieure	**estimation**	borne inférieure	borne supérieure	**estimation**	borne inférieure	borne supérieure
0	**1**	1	1	**1**	1	1	**1**	1	1
1	**0.80**	0.78	0.82	**0.73**	0.68	0.78	**0.81**	0.80	0.83
2	**0.64**	0.61	0.67	**0.53**	0.46	0.60	**0.66**	0.63	0.69
3	**0.51**	0.48	0.54	**0.39**	0.31	0.47	**0.54**	0.50	0.58
4	**0.46**	0.42	0.50	**0.34**	0.24	0.44	**0.48**	0.44	0.53
5	**0.41**	0.36	0.45	**0.30**	0.19	0.41	**0.43**	0.38	0.48
6	**0.36**	0.31	0.41	**0.26**	0.15	0.38	**0.39**	0.33	0.44
7	**0.33**	0.27	0.38	**0.21**	0.10	0.34	**0.36**	0.29	0.42
8	**0.30**	0.24	0.36	**0.17**	0.07	0.30	**0.33**	0.26	0.40
9	**0.27**	0.21	0.33	**0.14**	0.05	0.27	**0.31**	0.23	0.38
10	**0.23**	0.17	0.29	**0.12**	0.03	0.25	**0.26**	0.19	0.33
11	**0.19**	0.14	0.25	**0.11**	0.02	0.24	**0.22**	0.15	0.29
12	**0.16**	0.11	0.22	**0.09**	0.02	0.23	**0.18**	0.12	0.25
13	**0.13**	0.08	0.19	**0.06**	0.01	0.18	**0.15**	0.10	0.22
14	**0.10**	0.06	0.16	**0.04**	0.00	0.13	**0.13**	0.07	0.19
15	**0.08**	0.05	0.13	**0.02**	0.00	0.10	**0.10**	0.06	0.16
16	**0.07**	0.04	0.11	**0.02**	0.00	0.10	**0.08**	0.04	0.13
17	**0.05**	0.03	0.09	**0.02**	0.00	0.10	**0.06**	0.03	0.11
18	**0.04**	0.02	0.08	**0.01**	0.00	0.09	**0.05**	0.02	0.09
19	**0.04**	0.02	0.07	**0.01**	0.00	0.08	**0.04**	0.02	0.08
20	**0.03**	0.01	0.06	**0.01**	0.00	0.07	**0.03**	0.01	0.06

Champ : France métropolitaine.
Source : Insee, *Enquête famille et logements* (*EFL*) 2011 ; Ined-Insee, *Étude des relations familiales et intergénérationnelles* (*Erfi*), vagues 1 et 3, 2005 et 2011.

Tableau A2-2

Fonctions de survie estimées pour les femmes, associées aux risques instantanés de base lorsque l'on considère un ensemble de covariables (cause d'entrée, diplôme ou catégorie sociale)

	Cause d'entrée			Diplôme			Catégorie sociale		
	estimation	borne inférieure	borne supérieure	estimation	borne inférieure	borne supérieure	estimation	borne inférieure	borne supérieure
0	**1.00**	1.00	1.00	**1.00**	1.00	1.00	**1.00**	1.00	1.00
1	**0.89**	0.87	0.90	**0.85**	0.87	0.90	**0.84**	0.87	0.90
2	**0.79**	0.76	0.81	**0.73**	0.76	0.81	**0.71**	0.76	0.81
3	**0.70**	0.67	0.73	**0.62**	0.67	0.73	**0.60**	0.67	0.73
4	**0.65**	0.61	0.69	**0.57**	0.61	0.69	**0.55**	0.61	0.69
5	**0.61**	0.57	0.65	**0.52**	0.57	0.65	**0.50**	0.57	0.65
6	**0.57**	0.52	0.62	**0.48**	0.52	0.62	**0.46**	0.52	0.62
7	**0.54**	0.48	0.60	**0.45**	0.48	0.60	**0.42**	0.48	0.60
8	**0.51**	0.45	0.57	**0.42**	0.45	0.57	**0.38**	0.45	0.57
9	**0.49**	0.42	0.55	**0.40**	0.42	0.55	**0.34**	0.42	0.55
10	**0.44**	0.37	0.51	**0.35**	0.37	0.51	**0.30**	0.37	0.51
11	**0.40**	0.32	0.47	**0.31**	0.32	0.47	**0.26**	0.32	0.47
12	**0.36**	0.29	0.44	**0.27**	0.29	0.44	**0.23**	0.29	0.44
13	**0.32**	0.24	0.39	**0.23**	0.24	0.39	**0.19**	0.24	0.39
14	**0.28**	0.20	0.35	**0.20**	0.20	0.35	**0.16**	0.20	0.35
15	**0.24**	0.17	0.32	**0.17**	0.17	0.32	**0.13**	0.17	0.32
16	**0.21**	0.14	0.28	**0.14**	0.14	0.28	**0.10**	0.14	0.28
17	**0.18**	0.12	0.25	**0.11**	0.12	0.25	**0.08**	0.12	0.25
18	**0.15**	0.09	0.22	**0.09**	0.09	0.22	**0.06**	0.09	0.22
19	**0.13**	0.08	0.19	**0.08**	0.08	0.19	**0.05**	0.08	0.19
20	**0.11**	0.07	0.16	**0.07**	0.07	0.16	**0.04**	0.07	0.16

Champ : femmes, France métropolitaine.
Source : Insee, *Enquête famille et logements* (*EFL*) 2011 ; Ined-Insee, *Étude des relations familiales et intergénérationnelles* (*Erfi*), vagues 1 et 3, 2005 et 2011.

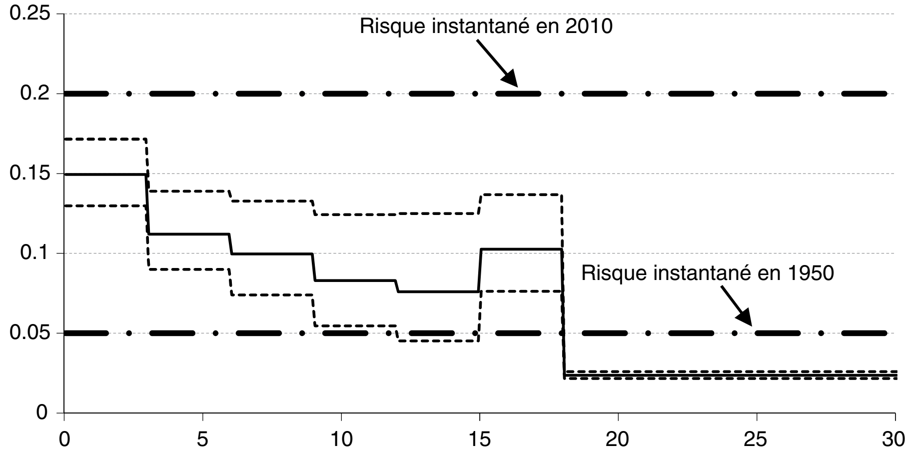
ANNEXE 3

SIMULATION DES CORRÉLATIONS ENTRE D ET T

Pour tenter de saisir ce qui se passe lorsque l'on relâche l'hypothèse d'indépendance entre la variable de flux *D* et la variable de durée *T*, on simule une population de 100 000 personnes entrant en famille monoparentale de manière uniforme entre 1950 et 2010, mais dont la durée est négativement corrélée à la date d'entrée. Pour cela, on simule *T* selon une loi de Weibull de paramètre de forme constant qui vaut 1 (cas particulier pour lequel la loi est exponentielle) et un paramètre d'échelle qui vaut a-b.*D*, où a=507.5 et b=0.25. Ainsi, pour les personnes entrées en monoparentalité en 1950, la durée moyenne est de 20 ans tandis pour celles entrées en 2010, la durée moyenne est de 5 ans (bien sûr cette situation est exagérée et ne correspond pas à la réalité). Sur la population simulée, la durée moyenne est de 12.4 ans.

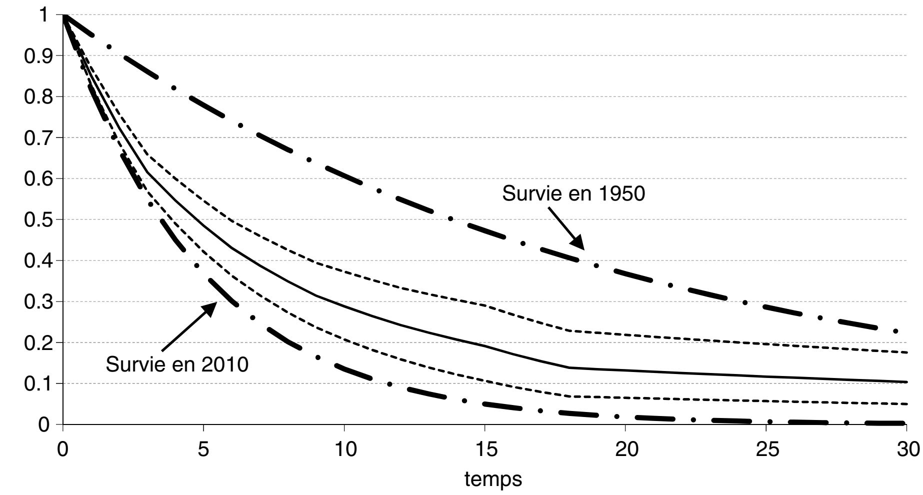
La durée moyenne estimée est de 8.2 ans. On a donc tendance à sous-estimer les durées lorsqu'il y a une corrélation négative entre la durée *T* et la date d'entrée en famille monoparentale *D*. Néanmoins, les estimations donnent bien un risque instantané et une survie qui se situent entre les risques instantanés et survies extrêmes, à savoir ceux de 1950 et de 2010 (figures A3-I et A3-II).

Figure A3-I
Estimation du risque instantané en présence d'une corrélation entre la variable D et la variable T



Note : on a simulé pour 100 000 personnes des durées dont la loi dépend de la date D d'entrée en famille monoparentale selon une loi de Weibull(1,a-b.D) où a et b sont des coefficients valant 507.5 et 0.25. Les traits en pointillés représentent l'intervalle de confiance à 95 % obtenu par la méthode du maximum de vraisemblance.

Figure A3-II
Estimation de la survie en présence d'une corrélation entre la variable D et la variable T



Note : on a simulé pour 100 000 personnes des durées dont la loi dépend de la date D d'entrée en famille monoparentale selon une loi de Weibull(1,a-b.D) où a et b sont des coefficients valant 507.5 et 0.25. Les traits en petits pointillés représentent l'intervalle de confiance à 95 % obtenu par la méthode du maximum de vraisemblance.

BIAIS DE L'ESTIMATEUR DE LA MOYENNE

Supposons qu'on dispose d'une variable de durée continue notée T_c et d'une autre variable de durée discrète, ne prenant que des valeurs entières, notée T_d. Supposons de plus que la fonction de survie S_d de la variable discrète est égale, pour les durées à valeurs entières, à la fonction de survie S_c de la variable continue. Autrement dit, $P(T_d \geq t) = P(T_c \geq t) \forall t \in N$. Alors on montre que $E[T_d] \approx E[T_c] - 0{,}5$.

$$E[Td] = \sum_{u\geq 1} S_d(u) = \sum_{u\geq 1} S_c(u) = \sum_{u\geq 1} [\int_{u-1}^{u} S_c(x).dx - \int_{u-1}^{u} (S_c(x) - S_c(u)).dx] = \sum_{u\geq 1}\int_{u-1}^{u} S_c(x).dx - \sum_{u\geq 1} R(u) = E[T_c] - R$$

où $R = \sum_{u\geq 1} R(u)$ et $R(u) = \int_{u-1}^{u} (S_c(x) - S_c(u)).dx$.

Si on approxime $S_c(x)$ par une fonction linéaire entre u-1 et u, alors on a $S_c(x) - S_c(u) \approx (S_c(u-1) - S_c(u))(u-x)$ et donc $R(u) \approx 0{,}5(S_c(u) - S_c(u-1))$. D'où $R = \sum_{u\geq 1} R(u) \approx 0.5$.

On en conclut que $E[T_d] \approx E[T_c] - 0{,}5$.